UNIVERSITY OF CALIFORNIA, SAN DIEGO

A Study of the Standard Model Higgs, WW and ZZ Production in Dilepton plus Missing Transverse Energy Final State at CDF Run II

A dissertation submitted in partial satisfaction of the

requirements for the degree

Doctor of Philosophy

in

Physics

by

Shih-Chieh Hsu

Committee in charge:

Professor Frank Würthwein, Chair
Professor Jim Branson
Professor Claudio Campagnari
Professor Philip E. Gill
Professor Benjamin Grinstein
Professor Rainer Wallny

2008

The dissertation of Shih-Chieh Hsu is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2008

To Rong-Yuang Hsu and Yu-Yin Liu.

EPIGRAPH

The mind is not a vessel to be filled but a fire to be kindled

—Plutarch (45-127 AD)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This thesis summarizes the knowledge I have learned and the frontiers of physics I have explored as a graduate student working with the UCSD High Energy Physics group. I would like to thank many people who supported and accompanied me along the road to understand how nature works. This thesis is devoted to my parents who raised me with endless patience and love. It took some efforts but I am appreciated that you have eventually understood my decision to study abroad.

My early days in UCSD trace back to 2003. Professor David MacFarlane opened the door for me to come to UCSD and gave me the freedom to find my research interests. Professor Hans Paar trained me in hardware during summer 2004 at BaBar, including the physics of sailing, and taught me tips to identify a good restaurant - a real restaurant always use a real table cloth. There is one thing I should apologize to Hans for is that I still have not tried real sailing.

Professor Frank Würthwein, my respected advisor, your wisdom and the talent and insight to see through my problems and judge the appropriate direction to follow has become my best model to pursue my career. Your guidance inspired me in the creative way of understanding and thinking about physics on my feet. You always have positive attitudes and bright energies to push things forward. Your thoughts to ensure my education resources maintain the best quality. I thank you so much for establishing a great group at CDF for me to work with. I will apply your skills to figure out the most valuable treasures from the whole ginbang.

Professor Elliot Lipeles, the most clever physicist I have ever met, directed me to work on the diboson physics program at CDF. You trained me in all the computing knowledge required to operate the CAF and the skills necessary to develop JobMon. You showed me how to do analysis and proposed the great idea of applying the matrix element method to the Higgs search which is the core part of this thesis. New ideas keep popping out from your brain just like endless running water. I will keep telling myself "Yes! There is always a smarter way to do things."

Professor Mark Neubauer, the most reliable partners, mentored me in how to push things through and finish projects under a tight schedule and pressure without going panic. The depth and breadth of your knowledge and the ability to overcome problems without loosing focus taught me not to be afraid of making changes and organizing things better to turn an idea into reality. Neutrino physics, B physics and Diboson physics

- it's always amazed to see you switch to different subjects and establish important contributions. I wish I have been an good officemate to you for three years and I'll keep staying away from hamburgers everyday in my mind. Don't be surprised if you hear people talk about your "sea-saw learning curve theory". I have started advertised your philosophy of education within my friends.

Mattew Norman, my good friend since I arrived UCSD, has assisted me a lot to get used to USA culture and learned physics together. I remember how we argued physics, history and those evening walks from Mayer Hall to the apartment. Rami Vanguri is a good partner to work on analysis together and also taught me American culture as well as how to speak English. I can now order olives correctly but still disagree that Bolo Yeung and I look like twins.

There are many good friends at CDF that helped me to move on this far. Anadi Canepa and Else Lytken, helped familiarize me with the CDF analysis framework and educate me in SUSY physics. Pasha Murat and Ray Culbertson led efforts to develop and produce Stntuple, the stage previous to the Diboson ntuple. Daniel Whiteson shared tips to speed up the event probability calculations and excited me with his ideas of multivariate analysis. Chris Hays and David Waters helped me to build up the basis of understanding leptons and diboson physics analysis at CDF. I should also thank the Electroweak and Higgs group conveners for your professional opinions and constructive support in pushing forward the analysis. I would also like to thank all the diboson group, the HWW group and the CDF collaborations.

Professor Jim Branson, Professor Vivek Sharma and Professor Avi Yagil. You completed my Ph.D. training when I moved back to San Diego in Fall 2007. I gained from you the ability to show confidence and be aggressive. I learned how to think from the bottom-up. The size of the face of the EM tower, 24 cm x 48 cm, has been ironed into my memory. I like to thank Elizabeth Dusinberre and Jon Spalding to proof read and make my thesis readable.

Here's special thanks to my good friends: Shin-Shan (Eiko) Yu, Rong-Shyang Lu, Shan-Hwei (Sunny) Chung and Song-Ming Wang. You colored my life with laughs, jokes, stories and shared with your life experiences, also rich knowledge of CDF and analysis. Shang-Lun Yeh, my love, you supported me going this far. Our sweet daily talks cross ten thousand miles distance is the fuel to drive me on. This thesis also celebrates our marriage! What a wonderful year.

VITA

2003-2008	**Ph. D. in Physics**, University of California San Diego, La Jolla, CA 92093 Thesis Adviser: Prof. Frank Würthwein Thesis Title: *A Study of the Standard Model Higgs, WW and ZZ Production in Dilepton plus Missing Transverse Energy Final State at CDF Run II*
2001-2003	**Second Lieutenant in Taiwanese Navy** , Tsuoin Navy Base, Kaohsiung, Taiwan
1999-2001	**Master of Science in Physics** , National Taiwan University, Taipei, Taiwan Thesis Advisor: Prof. Paoti Chang Thesis Title: *Study of the Charmless Hadronic Decay $B \to \eta' K$ with the Belle Detector*
1995-1999	**Bachelor of Science in Physics** , National Taiwan University, Taipei, Taiwan

PUBLICATIONS

T. Aaltonen, *et al.* (CDF Collaboration) "Search for a Higgs Boson Decaying to Two W Bosons at CDF," *hep-ex/arXiv:0809.393, Submitted to Phys. Rev. Lett on Sep 23 2008* **(this thesis)**

T. Aaltonen, *et al.* (CDF Collaboration) "Strong Evidence for ZZ Production in $p\bar{p}$ Collisions at $\sqrt{s}$=1.96 TeV," *Phys. Rev. Lett. 100, 201801 (2008).* **(this thesis)**

S.-C. Hsu (for the CDF Collaboration), "Searching for $H \to WW^{(*)}$ and Other Diboson Final States at CDF," *hep-ex/arXiv:0706.2200.*

A. Abulencia, *et al.* (CDF Collaboration), "Observation of WZ Production," *Phys. Rev. Lett. 98, 161801 (2007).*

K. Abe, *et al.* (Belle Collaboration), "Measurement of the Branching Fraction for $B \to \eta' K$ and Search for $B \to \eta' \pi^+$," *Phys.Lett. B517 (2001) 309-318*

C. Steenberg, **S.-C.Hsu**, E. Lipeles, F. Würthwein, "JobMon: A Secure, Scalable, Interactive Grid Job Monitor," *Proceedings of Computing in High-energy Physics (CHEP 06), Mumbai, India, (2006).*

M. Casarsa, **S.-C. Hsu**, E. Lipeles, M. Neubauer, S. Sarkar, I. Sfiligoi and F. Würthwein, "The CDF Analysis Farm," *AIP Conf. Proc.* **794**, *275 (2005).*

ABSTRACT OF THE DISSERTATION

A Study of the Standard Model Higgs, WW and ZZ Production in Dilepton plus Missing Transverse Energy Final State at CDF Run II

by

Shih-Chieh Hsu

Doctor of Philosophy in Physics

University of California San Diego, 2008

Professor Frank Würthwein, Chair

We report on a search for Standard Model (SM) production of Higgs to WW^* in the two charged lepton (e, μ) and two neutrino final state in $p\overline{p}$ collisions at a center of mass energy $\sqrt{s} = 1.96$ TeV. The data were collected with the CDF II detector at the Fermilab Tevatron and correspond to an integrated luminosity of 1.9fb^{-1}. The Matrix Element method is developed to calculate the event probability and to construct a likelihood ratio discriminator. There are 522 candidates observed with an expectation of 513±41 background events and 7.8±0.6 signal events for Higgs mass 160GeV/c^2 at next-to-next-to-leading logarithmic level calculation. The observed 95% C.L. upper limit is 0.8 pb which is 2.0 times the SM prediction while the median expected limit is $3.1^{+1.3}_{-0.9}$ with systematics included. Results for 9 other Higgs mass hypotheses ranging from 110GeV/c^2 to 200GeV/c^2 are also presented.

The same dilepton plus large transverse energy imbalance ($\not\!\!E_T$) final state is used in the SM ZZ production search and the WW production study. The observed significance of $ZZ \to ll\nu\nu$ channel is 1.2σ. It adds extra significance to the $ZZ \to 4l$ channel and leads to a strong evidence of ZZ production with 4.4 σ significance. The potential improvement of the anomalous triple gauge coupling measurement by using the Matrix Element method in WW production is also studied.

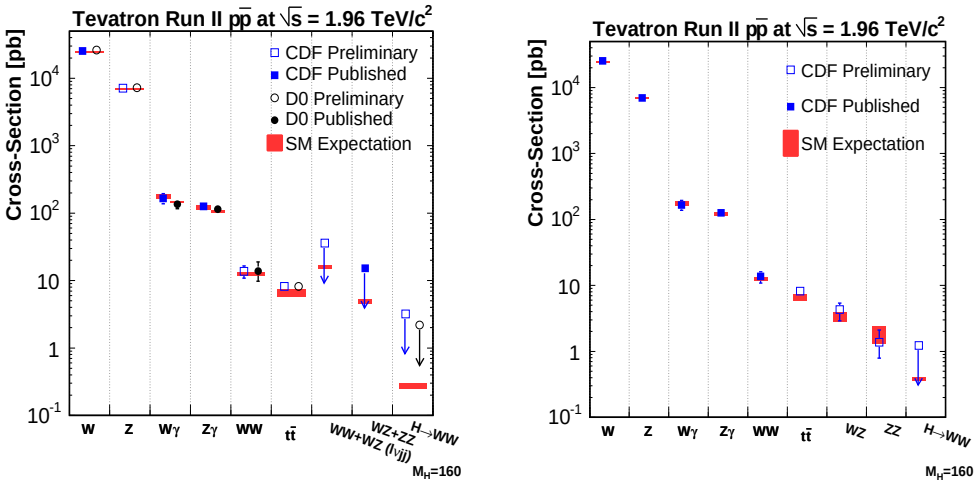
Chapter 1

Overview

In this dissertation, we study the electroweak spontaneous symmetry breaking and gauge structure of the Standard Model (SM) by searching for a scalar Higgs boson and testing the diboson productions ZZ and WW. The Standard Model has triumphantly explained the fundamental building blocks of matter and their interactions since 1972. Every component of the Standard Model has been found after the discovery of the Top quark in 1995. One of the big questions to be verified in the Standard Model is how electroweak symmetry breaking proceeds. A single scalar Higgs boson is the consequence of the minimal way to break electroweak symmetry through Higgs mechanism.

The consequences are that all fundamental particles acquire masses and the WW scattering amplitude is unitaritrized. The search for the Standard Model Higgs boson is therefore one of the crucial research topics in modern particle physics. Other mechanisms could generate symmetry breaking and predict Higgs-like particles, e.g. Supersymmetry. It is also possible to generate symmetry breaking without the Higgs, e.g. strong dynamic symmetry breaking or Technicolor. One of the most general ways to explore new physics is to study the self-coupling between gauge bosons, the mediator of the interactions. A thorough understanding of diboson productions and precision measurements of triple gauge boson couplings may reveal new phenomena beyond the Standard Model [1]. Diboson production has been extensively tested in the e^+e^- collider LEP. The WW cross-section is known [2] and ZZ production has been observed [3]. However, $p\bar{p}$ collisions at Tevatron probe large center-of-mass energy region which is exclusive to LEP.

Figure 1.1(a) shows the status of experimental observations and theoretical



(a) Summer 2006 (b) June 2008

Figure 1.1: The theoretical production cross-section and experimental status of QCD or electroweak processes at the Tevatron in summer 2006 (a) and the results from this thesis in June 2008 (b). The theoretical cross-sections are calculated to next-to-leading order (QCD). Each arrow line indicates the 95% Confidence Level production cross-section limit.

calculations for physics processes relevant to this thesis in summer 2006. The 95% Confidence Level (C.L.) production cross-section limit of Higgs was factor of 12 larger than the Standard Model prediction [4]. The cross-section of WW production was measured [5] but there were no evidence of WZ and ZZ productions [6].

Figure 1.1(b) shows the new knowledge provided by this thesis at the completion time in June 2008. The Higgs production cross-section limit has been pushed down to factor of 2 larger than the Standard Model prediction at 160 GeV/c^2 [7]. WW cross-section measurement becomes systematics dominant. ZZ production significance is 4.4σ excess above backgrounds [8]. WZ observation is established by using the same lepton selections described in this thesis [9].

After submission of this thesis, we continue to work on the Higgs search and set a new limit of 1.7 times larger than the Standard Model prediction at 160 GeV/c^2 with 3 [10].

We are looking for fully leptonic decay of $H \rightarrow WW^* \rightarrow ll\nu\nu$, $WW \rightarrow ll\nu\nu$ and $ZZ \rightarrow ll\nu\nu$. The small branching ratio of $W \rightarrow l\nu$ (10%) and $Z \rightarrow ll$(3%) makes the production rarer. However, the signal to background ratio is 1000 times better than semileptonic decay where one of the W decays hadronically. Additionally, the neutrinos

have no interaction with the detectors and leave a large imbalance of transverse energy ($\not\!\!E_T$). One goal of this thesis is to develop a multivariate analysis technique to utilize all available information from the dilepton plus $\not\!\!E_T$ signature.

The analysis strategy is to maximize the acceptance of the event selection and then apply a multivariant technique to discriminate signal from background. The lepton types used are electrons or muons including ones decayed from taus. An advanced analysis technique, the Matrix Element method, is developed to exploit all the kinematic information by calculating the event probability from the theoretical production cross-section. Currently, the Fermilab Tevatron is the only facility which produces a large sample of diboson events. The next energy frontier, Large Hadron Collider, will begin operations in 2008. This dissertation can provide a guideline in the search for the Standard Model Higgs of the LHC in the $ll + \not\!\!E_T$ final state.

High Energy Physics is a field requiring big collaboration efforts and I would like to highlight the people who contributed explicitly to this thesis. I generated Monte Carlo sample, made the first round of lepton identification and event selection studies, implemented the event probability calculation using the Matrix Element method, and configured the code to do the statistical interpretation. Elliot Lipeles and Mark Neubauer expanded lepton categories and re-wrapped my low level objects in a new analysis framework - the Diboson package. Mark produced the analysis ntuples, implemented the event selection code and managed code versions to make sure the results are reproducible. Elliot measured lepton identification efficiencies and lepton mis-identification rates originated from jets (fake rate) and performed systematics studies. Frank Würthwein planned the strategy, examined the contents, plotted the direction of analysis and managed the social communications within the CDF group.

The organization of this thesis is as follows: Chapter 2 gives an overview of the Standard Model, the Higgs mechanism, new physics and triple gauge coupling formalism. Chapter 3 introduces the Tevatron accelerator facility, the CDF II detector and the CDF computing. Chapter 4 describes lepton identification, major background sources and event selection. Chapter 5 talks about our data modeling and demonstrates our procedures to analyze data. Chapter 6 explains Matrix Element method which utilizez all the leptonic kinematic information. Chapter 7 studies various systematics uncertainty and estimates the impacts to our results. Chapter 8, 9 and 10 report and discuss the results of the Higgs search, ZZ production and WW measurements, respectively. One

additional section 8.3.2 is added in Sep 2008. It is a follow-up of Higgs search using the technique of this thesis in 3 fb^{-1} and submitted to . Finally, Chapter 11 gives a summary of this thesis.

Chapter 2

Higgs and Diboson Physics in Hadron Collider

The standard model (SM) is currently the best knowledge we have to describe nature at the sub-atomic level. An introduction of the SM is presented in section 2.1. We will learn the definition and the importance of the gauge bosons, the Higgs, and hadrons. Section 2.2 further discusses the properties and knowledge frontier of the SM Higgs. We will review three new physics models which can enhance the Higgs production cross-section. Section 2.3 presents a general way to test the validity of SM via triple-gauge-boson coupling in diboson production. Section 2.4 talks about hadron collision physics. It introduces the concept of parton which makes the theoretical calculation in hadron collisions feasible. This chapter only provides minimal knowledge to appreciate the motivation of this thesis. More comprehensive treatments of the materials can be found in several review articles or textbooks, e.g. [11, 12].

2.1 Introduction to the Standard Model

The Standard Model (SM), built in the 20th century, is a description of the fundamental structure of the physical world. It requires twelve matter particles and four force carriers, summarizes three fundamental theories and successfully explains most of the experimental results. The three theories that describe the interactions are unified into two in the energy scale $O(10^2$ GeV), a.k.a electroweak energy scale. This unification still needs two independent coupling parameters and motivates physicists to keep looking for a unified theory using a single free parameter at a grand unification scale of

$O(10^{16}$ GeV) [13].

2.1.1 The Composition of the World

The twelve matter particles are categorized in two types - quarks and leptons. Matter particles are called fermions because their intrinsic angular momentum, spin, is half integer. They are point-like particles which means there are no more internal structures. There are six types (flavors) of quarks, which are grouped in three pairs: up (u)/down (d), charm (c)/strange(s), and top (t)/bottom (b). Each quark has mass and the up-type quark (up, charm, top) has a fractional charge 2/3 while down-type quark has a fractional charge (-1/3). Each group is called a generation of quarks. There are then six leptons, three with a negative charge and a mass - electron (e^-), muon (μ^-) and tauon (τ^-) - and three neutral and with very small mass - electron-neutrino (ν_e), muon-neutrino (ν_μ) and tau-neutrino (ν_τ). Each group of lepton and lepton-neutrino is also called a generation of leptons. The existence of neutrino mass was revealed in 1995. The masses of the leptons and quarks are listed in Table 2.1. We will treat neutrinos as massless particles due to the scale of the energy considered in this thesis, O(10 100 GeV), is much larger than neutrino masses.

Table 2.1: Three generations of leptons and quarks in the Standard Model and their masses [1].

Generation	Fermions		Mass [GeV/c^2]
I	electron neutrino	ν_e	$<$ 2.3 eV (95% C.L.)
	electron	e	0.00051099892(4)
	up quark	u	0.0015 to 0.003
	down quark	d	0.003 to 0.007
II	muon neutrino	ν_μ	$<$ 0.19 MeV (90% C.L.)
	muon	μ	0.105658369(9)
	charm quark	c	1.25 $\pm$ 0.09
	strange quark	s	0.095 $\pm$ 0.025
III	tau neutrino	ν_τ	$<$ 18.2 MeV (95% C.L.)
	tau	τ	1.77699(29)
	top quark	t	174.2 $\pm$ 3.3
	bottom quark	b	4.20 $\pm$ 0.07

For each particle, there is an associated antiparticle with the same mass and opposite charge. For example, the anti-particle of the electron is the positron (e^+) and the anti-particle of the up quark is the anti-up quark ($\bar{u}$) pronounced as u-bar. Quarks

are the main constituents of composite particles called hadrons. There are two types of hadrons: mesons are composed of a quark and an anti-quark, e.g. pions π ($u\bar{d}$), kaons K($u\bar{s}$); baryons are composed of three quarks, e.g. protons p (uud) and neutrons n (udd). The fractional charges of the quarks are hidden in the charges of the hadrons which are all integers

There are three types of interactions describing the behavior of the particles and how particles compose compounds. The free quarks are confined in the hadrons through strong interactions. The same interaction binds protons and neutrons to form nuclei. Nuclei and electrons can compose an atom through electromagnetic (EM) interactions. EM interactions between several atoms can form a molecule and molecules further build up everything in the world. There is another type of interaction, weak interaction, describing the phenomenon of radiative decay, e.g. beta decay. Quarks are the only matter particles that can experience all three interactions. Charged leptons only have weak and EM interactions. Neutrinos only experience the weak interaction.

These three interactions, described as communication between particles by the exchange of special"force-carrying particles" called bosons, have spin integer and carry energy from one particle to another. Each force has its own characteristic bosons: the gluon (strong force), the photon (electromagnetic force), and the $W^{\pm}$ and Z bosons (weak force). Table 2.2 lists the properties of the force carriers.

Table 2.2: Force carriers in the Standard Model and their masses [1].

Force	Bosons		Mass [GeV/c^2]
electromagnetic	photon	γ	0
charged weak	W boson	$W^{\pm}$	80.403 $\pm$ 0.029
neutral weak	Z boson	Z^0	91.1876 $\pm$ 0.0021
strong	gluon	g	0

The theories of the three forces were highlighted by a series of Nobel prizes. Quantum electrodynamics (QED), which describes the interactions of all charged particles, resulted in the 1965 Nobel prize to Sinitiro Tomonaga, Julian Schwinger and Richard P. Feynman. The weak interactions were developed to describe the underlying theory of beta decay and allow flavor changing between the fermions. A big success of the Standard Model is the unification of the electromagnetic and the weak forces into the so-called electroweak force. Sheldon L. Glashow, Abdus Salam and Steven Weinberg were awarded the 1974 Nobel prizes for developing of electroweak unification theory [14, 15].

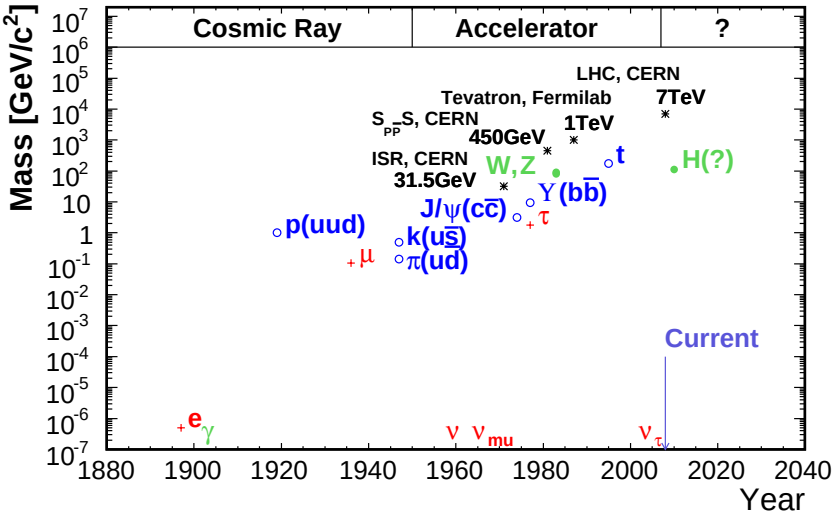


Figure 2.1: The timeline for discoveries in particle physics and proton colliders. Colliders become the dominant way to search for particles after 1950.

Gerardus'T Hooft, and Martinus J.G. Veltman were awarded the 1999 Nobel prizes for elucidating the quantum structure of electroweak interactions in physics. However, this unification requires a spontaneous symmetry breaking for which the minimal mechanism is the Higgs mechanism. It predicts one extra Higgs field which provides the mass to all the particles and bosons and a corresponding neutral Higgs particle which is yet to be found. We will discuss the electroweak theory and Higgs mechanism in Section 2.2.

Quantum chromodynamics [16] is the theory which describes the strong interactions between quarks. It resulted in the 2004 Nobel Prizes, awarded to David J. Gross, H. David Politzer and Frank Wilczek for the discovery of asymptotic freedom in the theory of the strong interaction. The coupling strength of the forces depends on the energy scale. At the GUT energy scale the strength of the three interactions are approximately the same, which leads to a speculation that there is a grand unification of the three forces at that scale. A possible unification of gravity with the other three interactions is more mysterious as it is not even included in the standard model. One of the best candidates for explaining this is called string theory which may one day unify all forces in higher dimensions. The scope of this thesis will focus on the examination of the electroweak theory.

Figure 2.1 shows the time-line for discoveries in particle physics [11]. The period before 1950 was called cosmic ray era, because of new particles mostly relied on

naturally occuring high energy particles from space called cosmic rays. The invention of accelerators opened a new window to study particle physics and remains the major way to discover new particles. The heavier the particles the higher the beam energies are required. The first hadron collider, Intersecting Storage Ring (ISR), is built in 1971 with beam energy 31.5 GeV in collision with proton-proton. Tevatron is the current energy frontier with center-of-mass energy $\sqrt{s} = 1.96$ TeV. It contributs to the top quark discovery in 1995 and will continue to deliver the best knowledge concerning the Higgs boson and the existence of new physics until the startup of the Large Hadron Collier ($\sqrt{s} = 14$ TeV) in 2008. What is the next energy frontier we can explore after LHC? We need new ideas to study fundamental particle physics in order to not constraint ourselves by the limit of earth-based accelerators.

2.1.2 Spontaneous Symmetry Breaking via the Higgs Mechanism

Spontaneous symmetry breaking in physics takes place when a system that is symmetric with respect to some symmetry group enters a ground state that is not symmetric. At this point the system no longer behaves in a symmetric manner. The Higgs mechanism is the simplest known mechanism to provide spontaneous symmetry breaking in electroweak theory and breaks the unified theory into QED and Weak interactions at energies below O(1TeV). It predicts a new massive scalar boson and explains the origin of masses of every particle. A brief description of the mathematical formalism of the SM is shown in this section. Only key equations are shown to underline the important physics concepts. Detailed derivations and interpretation can be found in [12, 17].

The mathematical formalism of the standard model is formulated as relativistic quantum field theories. A quantum field is a quantum mechanical system containing a large, and possibly infinite, number of degrees of freedom. A field is an assignment of a physical quantity to every space-time point. One can describe matter particles and force carriers as fermion fields and gauge fields, whose interactions mandate the laws of physics. The fields are constructed from a symmetry group which satisfy symmetry group transformation. If the gauge (phase) transformation of the field preserve the original form of interactions, we then call the field theory *gauged*. The theories in standard model are all quantum gauge field theories.

We start categorizing matter fields Ψ by the group $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$. QCD is based on the $SU(3)_c$ group, with eight color charges. $SU(2)_L \otimes U(1)_Y$ is the

group of electroweak theory which is labeled by two conserved quantum numbers: weak isospin T^a and hypercharge Y. The matter fields are described with the three generations of left-handed and right-handed chiral quarks and leptons, $f_{L,R} = \frac{1}{2}(1 \mp \gamma_5)f$. The left-handed fermions are in weak isodoublets, while the right-handed fermions are in weak isosinglets. We will assume that the neutrinos, are massless and appear only with their left-handed components. The matter fields Ψ are listed in detail below:

$$\begin{aligned} L &= \left(\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L \right) \qquad R = (e_R^-, \mu_R^-, \tau_R^-) \\ Q &= \left(\begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L \right) \qquad \begin{aligned} U_R &= (u_R, c_R, t_R) \\ D_R &= (d_R, s_R, b_R) \end{aligned} \end{aligned} \tag{2.1}$$

There are gauge fields corresponding to the spin-one bosons that mediate the interactions. In QCD, the gluon field G_μ corresponds to the color field of the $SU(3)_C$ group. In the electroweak sector, we have the three fields $W_\mu^{1,2,3}$ which correspond to the generators T^a [with a=1,2,3] of the $\mathrm{SU(2)_L}$ group and the gauge field B_μ which corresponds to the generator Y of the $\mathrm{U(1)_Y}$ group. The field strengths are given by

$$\begin{aligned} G^a_{\mu\nu} &= \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s\, f^{abc} G^b_\mu G^c_\nu \\ W^a_{\mu\nu} &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g_2\, \epsilon^{abc} W^b_\mu W^c_\nu \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned} \tag{2.2}$$

where g_s and g_2 are, respectively, the coupling constants of $\mathrm{SU(3)_C}$ and $\mathrm{SU(2)_L}$.

The matter fields Ψ are minimally coupled to the gauge fields through the covariant derivative D_μ which, in the case of leptons, is defined as

$$D_\mu \Psi = \left(\partial_\mu - i g_2 T_a W^a_\mu - i g_1 \frac{Y_q}{2} B_\mu \right) \Psi \tag{2.3}$$

where g_1 is the coupling constant of $\mathrm{U(1)_Y}$. Now that we have all the components to write down the gauge invariant standard model Lagrangian which summarizes the dynamics of the system:

$$\mathcal{L}_{\rm SM} = -\frac{1}{4}W^a_{\mu\nu}W_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G^a_{\mu\nu}G_a^{\mu\nu} \quad \left\{ \begin{array}{l} W^{\pm}, Z, \gamma, \text{gluon kinetic} \\ \text{energies and} \\ \text{self-interactions} \end{array} \right. \tag{2.4}$$

$$\begin{array}{ll} +\bar{L}_i\, iD_\mu\gamma^\mu\, L_i & \left\{ \begin{array}{l} \text{lepton kinetic energies} \\ \text{and their interactions} \\ \text{with } W^{\pm}, Z, \gamma \end{array} \right. \\ +\bar{R}_i\, i\left(\partial_\mu - ig_1\frac{Y_R}{2}B_\mu\right)\gamma^\mu\, R_i & \end{array}$$

$$\begin{array}{ll} +\bar{Q}_i\, i(D_\mu - ig_sT_aG^a_\mu)\gamma^\mu\, Q_i & \left\{ \begin{array}{l} \text{quark kinetic energies} \\ \text{and their interactions} \\ \text{with } W^{\pm}, Z, \gamma \text{ and gluon} \end{array} \right. \\ +\bar{U}_{Ri}\, i(D_\mu - ig_sT_aG^a_\mu)\gamma^\mu\, U_{R_i} & \\ +\bar{D}_{Ri}\, i(D_\mu - ig_sT_aG^a_\mu)\gamma^\mu\, D_{R_i} & \end{array}$$

This Lagrangian is invariant under local $\mathrm{SU(3)_C \times SU(2)_L \times U(1)_Y}$ gauge transformations for fermion and gauge fields when both fermion and gauge fields are massless. It is true that gluons and photons are massless particles. However, quarks and leptons have mass and the W and Z bosons require mass to be O(100 GeV) scale. To satisfy the observed physics in the low energy regime, we look for a modification of the SM Lagrangian that preserves the electromagnetic symmetry to maintain charge conservation but break the original symmetry group as

$$SU(3)_c \otimes SU(2)_L \otimes U(1)_Y \rightarrow SU(3)_C \otimes U(1)_{\rm QED}\ ,$$

i.e. after the spontaneous symmetry breaking, the sub-group $U(1)_{\rm QED}$, of dimension 1, should remain as a symmetry of the vacuum.

The Higgs mechanism is a minimal way to provide spontaneous symmetry breaking and preserves the $U(1)_{QED}$ symmetry while predicting the properties of weak interactions. It is started by adding a new field Φ into the SM Lagrangian, without mass terms for fermions and gauge bosons:

$$\mathcal{L}_{\rm mass} = (D_\mu\Phi)^\dagger(D_\mu\Phi) \quad \left\{ W^{\pm}, Z, \gamma\text{-Higgs coupling} \right. \tag{2.5}$$

$$-\mu^2\Phi^\dagger\Phi - \lambda(\Phi^\dagger\Phi)^2 \quad \left\{ \begin{array}{l} \text{Higgs Potential} \\ \text{mass terms and quartic coupling} \end{array} \right.$$

$$-\lambda_{L_{ij}}\bar{L}_i\Phi R_j - \lambda_{L_{ij}}\bar{R}_j(\Phi^\dagger L_i) \quad \left\{ \text{Lepton-Higgs Yukawa coupling} \right.$$

$$\begin{array}{ll} -\lambda_{D_{ij}}\bar{Q}_i\Phi D_{R_j} - \lambda_{D_{ij}}\bar{D}_{R_j}(\Phi^\dagger Q_i) & \\ -\lambda_{U_{ij}}\bar{Q}_i\tilde{\Phi}U_{R_j} - \lambda_{U_{ij}}\bar{U}_{R_j}(\tilde{\Phi}^\dagger Q_i) & \left\{ \text{Quark-Higgs Yukawa coupling} \right. \end{array}$$

where $\tilde{\Phi} = -i\tau_2\Phi^*$ is an $SU(2)_L$ doublet and τ_2 is Pauli matrix. We need to generate masses for the three gauge bosons $W^\pm$ and Z but the photon should remain massless and QED must stay an exact symmetry. In getting mass, the $W^\pm$ and Z acquire one new polarization state each, so the Higgs field must have at least three degrees of freedom. The simplest choice is a complex $SU(2)$ doublet of scalar fields ϕ with hypercharge Y_Φ

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix}, \quad Y_\Phi = +1 \tag{2.6}$$

For $\mu^2 < 0$ and $\lambda > 0$ in the Higgs potential term, the neutral component of the doublet field Φ will develop a vacuum expectation value (vev)

$$\langle\, \Phi \,\rangle_0 \equiv \langle\, 0 \,|\, \Phi \,|\, 0 \,\rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \quad \text{with} \quad v = \sqrt{-\frac{\mu^2}{\lambda}} \tag{2.7}$$

We can then rewrite the field Φ in terms of four fields $\theta_{1,2,3}(x)$ and $H(x)$ at first order:

$$\Phi(x) = \begin{pmatrix} \theta_2 + i\theta_1 \\ \frac{1}{\sqrt{2}}(v + H) - i\theta_3 \end{pmatrix} = e^{i\theta_a(x)\tau^a(x)/2v} \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v + H(x)\,) \end{pmatrix} \tag{2.8}$$

We make a gauge transformation on this field to move to the unitary gauge:

$$\Phi(x) \to e^{-i\theta_a(x)\tau^a(x)/2v}\, \Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{2.9}$$

then fully expand the term $|D_\mu\Phi)|^2$ of the Lagrangian $\mathcal{L}_{mass}$:

$$\begin{aligned} |D_\mu\Phi)|^2 &= \left| \left(\partial_\mu - ig_2 \tfrac{\tau_a}{2} W_\mu^a - ig_1 \tfrac{1}{2} B_\mu \right) \Phi \right|^2 \\ &= \tfrac{1}{2} \left| \begin{pmatrix} \partial_\mu - \frac{i}{2}(g_2 W_\mu^3 + g_1 B_\mu) & -\frac{ig_2}{2}(W_\mu^1 - iW_\mu^2) \\ -\frac{ig_2}{2}(W_\mu^1 + iW_\mu^2) & \partial_\mu + \frac{i}{2}(g_2 W_\mu^3 - g_1 B_\mu) \end{pmatrix} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \right|^2 \\ &= \tfrac{1}{2}(\partial_\mu H)^2 + \tfrac{1}{8} g_2^2 (v + H)^2 |W_\mu^1 - iW_\mu^2|^2 + \tfrac{1}{8}(v + H)^2 |g_2 W_\mu^3 - g_1 B_\mu|^2 \\ &= \tfrac{1}{2}\partial_\mu H \partial^\mu H + \tfrac{1}{4} g_2^2 (v + H)^2 W_\mu^+ W^{-\mu} + \tfrac{1}{8}(v + H)^2 (g_2^2 + g_1^2) Z_\mu Z^\mu \end{aligned} \tag{2.10}$$

Here we have re-defined the electroweak field to physical fields $W_\mu^\pm$, Z_μ and A_μ.

$$W^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2)\ , \ Z_\mu = \frac{g_2 W_\mu^3 - g_1 B_\mu}{\sqrt{g_2^2 + g_1^2}}\ , \ A_\mu = \frac{g_1 W_\mu^3 + g_2 B_\mu}{\sqrt{g_2^2 + g_1^2}} \tag{2.11}$$

and pick up the terms which are bilinear in the fields $W^\pm, Z$:

$$M_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2} M_Z^2 Z_\mu Z^\mu \tag{2.12}$$

The W and Z bosons have acquired masses.

$$M_W = \frac{vg_2}{2}\ ,\ M_Z = \frac{1}{2}v\sqrt{g_2^2+g_1^2} = \frac{g_2 v}{2\cos\theta_W} = \frac{M_W}{\cos\theta_W}, \tag{2.13}$$

where $cos\theta_W = g_2/\sqrt{g_1^2+g_2^2}$ and θ_W is called Weinberg angle which describes the rotation of W^3 and B to physical field Z and A. The photon A is still massless since there is no corresponding bilinear term $A_\mu A^\mu$. Plugging equation (2.9) to the Higgs potential in $\mathcal{L}_{\text{mass}}$ (2.5)

$$-\frac{1}{2}(-2\mu^2)H^2 + \frac{1}{4}\mu^2 v^2(\frac{4}{v^3}H^3 + \frac{1}{v^4}H^4 - 1) \tag{2.14}$$

The second term predict the triple and quartic Higgs coupling and the first term identified the Higgs mass

$$m_H = \sqrt{-2\mu^2} = \sqrt{\frac{\lambda}{2}}v. \tag{2.15}$$

Thus, we have achieved the goal: by spontaneously breaking the symmetry $\mathrm{SU(2)_L}\times$ $\mathrm{U(1)_Y} \rightarrow \mathrm{U(1)_{QED}}$, three Goldstone bosons, θ_a, have been absorbed by the $W^\pm$ and Z bosons to form their longitudinal components and to get their masses. Since the $\mathrm{U(1)_{QED}}$ symmetry is still unbroken, the photon which is its generator, remains massless as it should be. One scalar neutral Higgs particle is predicted with mass undetermined.

We can also generate the fermion masses from the invariant Yukawa Lagrangian in $\mathcal{L}_{\text{mass}}$(2.5) by using the same scalar field Φ, with hypercharge Y=1, and the isodoublet $\tilde{\Phi} = i\tau_2\Phi^*$, which has hypercharge Y=-1 [12]. The diagonalization of the Yukawa (quark-Higgs) coupling generated mass matrix yields the CKM matrix which tranforms weak eigenstates to mass eigenstates and induces quark mixing. The Yukawa (lepton-Higgs) coupling is analogous to quark sectors.

From low energy phenomenon (e.g. μ-decay), we can relate Fermi constant G_F and weak coupling parameter g_2 to vev through,

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8m_W^2} = \frac{1}{2v^2} \tag{2.16}$$

Since G_F is measured accurately from muon lifetime, the vev is calculated to be

$$v = (\sqrt{2}G_F)^{-1/2} = 246\ \text{GeV}. \tag{2.17}$$

The W mass and Z mass are

$$m_W^2 = \frac{e^2}{4\sin\theta_W^2}v^2 = \frac{\pi\alpha}{\sin\theta_W^2}v^2 \sim (\frac{37.2}{\sin\theta_W}\ \text{GeV})^2 \sim (80\ \text{GeV})^2$$
$$m_Z^2 = (\frac{m_W}{\cos\theta_W})^2 \sim (90\ \text{GeV})^2. \tag{2.18}$$

2.2 Higgs Boson Properties

This section describes details of the Higgs boson properties. First we will investigate the Electroweak Lagrangian after spontaneous symmetry breaking (SSB). The coupling of the Higgs boson to gauge bosons and the fermions and thus the rates of decay for the Higgs boson into these particles are presented. Second we will discuss the production of Higgs boson at Tevatron. Third, we will mention the theoretical bounds of the Higgs mass and present our experimental knowledge of the Higgs bosons. Finally, we discuss the particular decay signature of $H \to WW^* \to l^+l^-\nu\bar{\nu}$ due to spin 0 properties of the SM Higgs boson.

2.2.1 Decay and Production

Decay

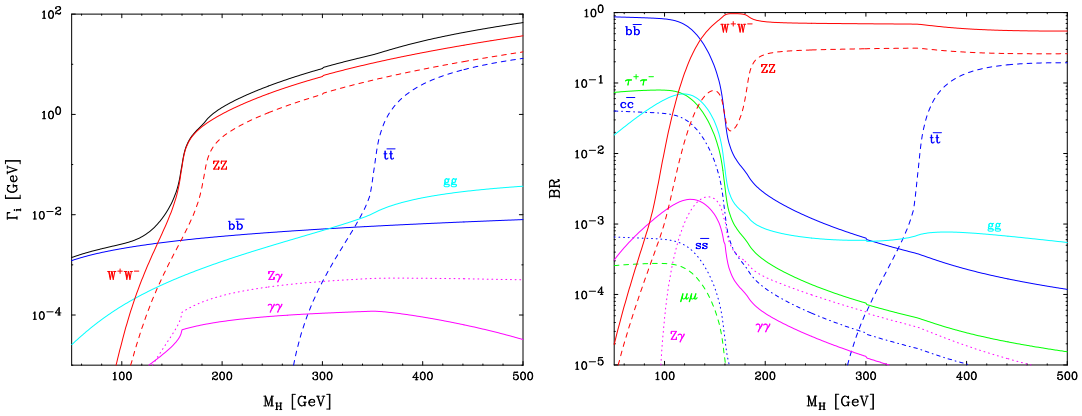
The Higgs boson couples to physical gauge bosons after SSB through covariant derivative of Higgs field $D_\mu\Phi$. Picking HWW and HZZ terms from Equation (2.10),

$$\begin{aligned} \frac{vg_2^2}{2} H W_\mu^+ W^{-\mu} &= g_2 M_W H W_\mu^+ W^{-\mu}, \\ \frac{v(g_1^2+g_2^2)}{4} H Z_\mu^+ Z^{-\mu} &= \frac{g_2 M_Z}{\cos\theta_W} H Z_\mu^+ Z^{-\mu}. \end{aligned} \tag{2.19}$$

The Higgs boson couples to fermions through the Yukawa coupling in the Lagrangian (2.5) with coupling strength m_f/v. Thus the Higgs can only decay directly to gauge boson pars or fermion pairs through a point interaction.

Figure 2.2 shows the partial widths of Higgs decaying to different final states as a function of m_H. The number is generated by HDECAY [18] in next-to-leading order calculation. Figure 2.2(b) shows the branching ratio of Higgs decay which is the ratio of partial width of each decay channel to total width. The coupling strengths of $H \to VV^*$ and $H \to f\bar{f}$ are proportional to m_H^3 and $m_f^2 m_H$, respectively. We might expect the partial width of vector boson pair final state is predominantly larger than the fermion pair final state. In reality, we have to take phase-space volume into account.

For $m_H < 2m_V$, one of the vector boson remains on-shell but the other vector boson is off-shell. The partial width is suppressed by the Breit-Wigner tail of off-shell vector boson. The partial width of fermion pair can be greater than than of vector boson pair for lower m_H. The heaviest quark pair, $b\bar{b}$, is the dominant decay channel.



(a) (b)

Figure 2.2: (a) Partial widths of the SM Higgs as a function of mass, m_H. The total width (sum of all partial widths) is the black curve [19]. (b) Branching ratios of the SM Higgs decaying to individual final states.

In the mass range between $2M_W$ and $2M_Z$, the partial widths of WW and ZZ increase rapidly. In the range of $2m_W \pm 2\Gamma_W$, where Γ_W is the Breit-Wigner width of W, WW becomes dominant decay channel. However, ZZ is still suppressed around this WW threshold point due to Breit-Wigner suppression. As a result, we see a dip of branching ratio of all the decay channels except WW because the WW partial width is increasing so rapidly near threshold. In the range $m_H > 2m_Z + 2\Gamma_Z$, ZZ has pass the threshold point and the decay width is half of the WW decay because of the identical particle quantum correction.

For $m_H > 2M_t$, the decay width of $H \to t\bar{t}$ becomes sizeable, but is still smaller than $\Gamma(H \to WW)$ and $\Gamma(H \to ZZ)$. A heavy Higgs becomes broader as m_H increases. At $m_H \sim 600$ GeV/c^2, the width is around 100 GeV; while for $m_H \sim 1$ TeV, Γ_H is already of the same size as the Higgs mass itself.

Higgs does not couple directly to massless particles: γ and gluon. It can decay to these final states via loop-induced transition as depicted in Figure 2.3. The same mechanism applies to Higgs to $Z\gamma$ and $\gamma\gamma$ decay. However, the loop particle can be a weak charged boson W. The destructive interference of fermion loop and W loop cause a weird partial width as shown in the Figure 2.2(a).

Production at Tevatron

In the Standard Model, the main production mechanisms for Higgs particles at hadron colliders makes use of the fact that the Higgs boson couples preferentially to heavy particles, that is the massive W and Z vector bosons, the top quark. The four main production processes at the Tevatron are thus: the gluon-gluon fusion mechanism: $gg \rightarrow H$ via top quark loop, the associated production with W/Z bosons: $q\bar{q} \rightarrow V + H$, the weak vector boson fusion processes: $qq \rightarrow V^*V^* \rightarrow qq + H$, and the associated Higgs production with heavy top or bottom, $gg, q\bar{q} \rightarrow Q\bar{Q} + H$. quarks. Figure 2.4 shows the leading order Feynman diagrams of Higgs production for each. The production cross-section is dominanted by the $gg \rightarrow H$ in a top-loop transition. Vector boson associated production is about one order of magnitude smaller. Vector boson fusion is smaller than associated vector boson production for $m_H < 140$ GeV/c^2. It becomes larger than vector boson associated production when $m_H > 190$ GeV/c^2. The Feynam diagram for the signal that is searched for in this thesis is shown in Figure 2.6.

2.2.2 Constraints

Theoretical Constraint

The upper bound of Higgs mass can be constrained by unitarity condition [11]. It is calculated from the amplitude of longitudinal vector boson pair scattering, $V_LV_L \rightarrow V_LV_L$, at high center-of-mass energy, where $V = W$ *or* Z. After imposing s-wave unitarity condition, we can get a Higgs mass upper bound

$$M_H \leq (\frac{8\pi\sqrt{2}}{3G_F})^{1/2} \approx 1TeV. \tag{2.20}$$

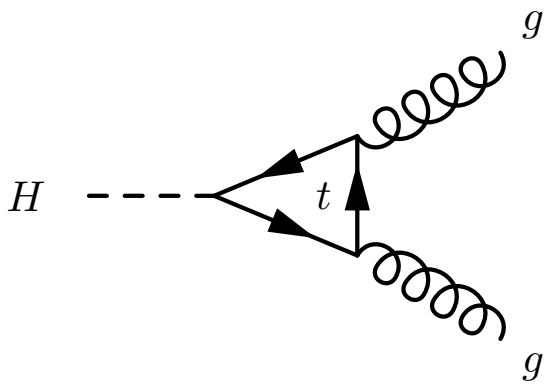


Figure 2.3: Feynman diagram for Higgs to gluon pair decay via loop-induced transition.

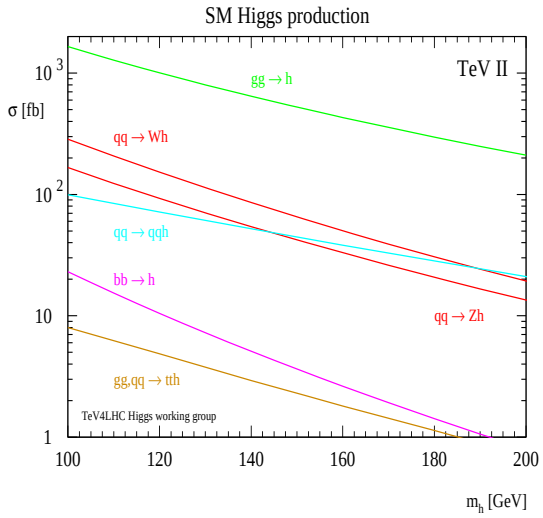


Figure 2.4: The production cross-section of the Standard Model Higgs boson at the Tevatron [19].

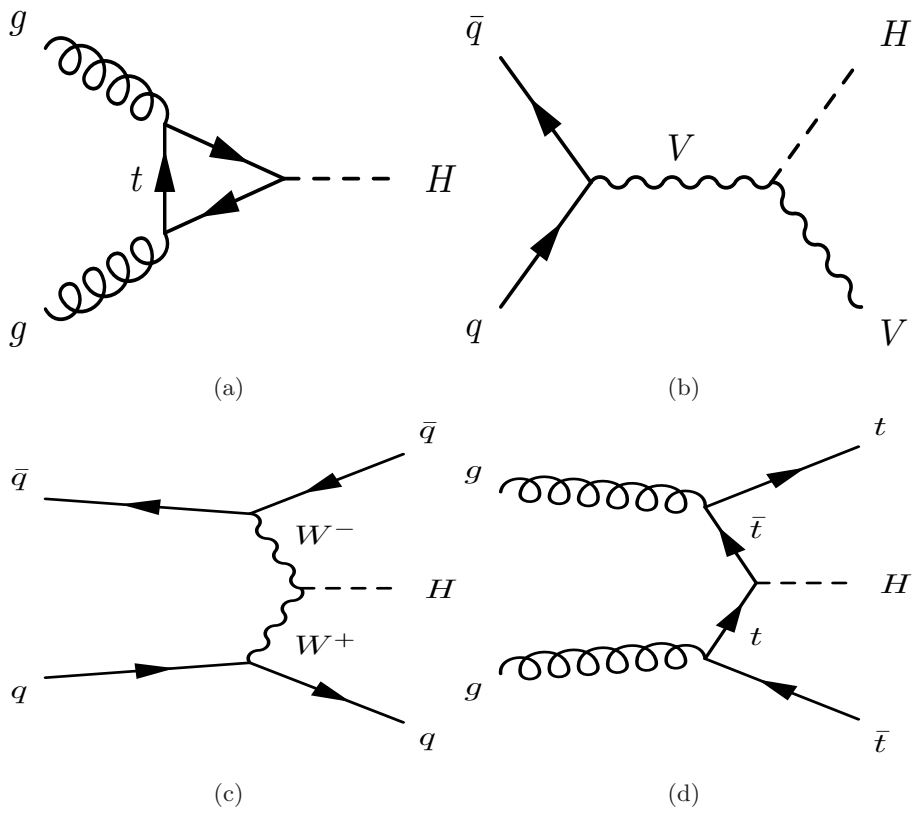


Figure 2.5: (a) gluon-gluon fusion. (b) associated production with vector boson. (c) vector boson fusion. (d) associated production with heavy quarks.

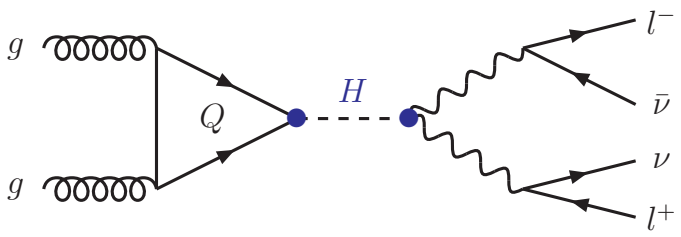


Figure 2.6: The Feynman diagram for the gg$\rightarrow H \rightarrow WW^* \rightarrow l^+l^-\nu\bar{\nu}$.

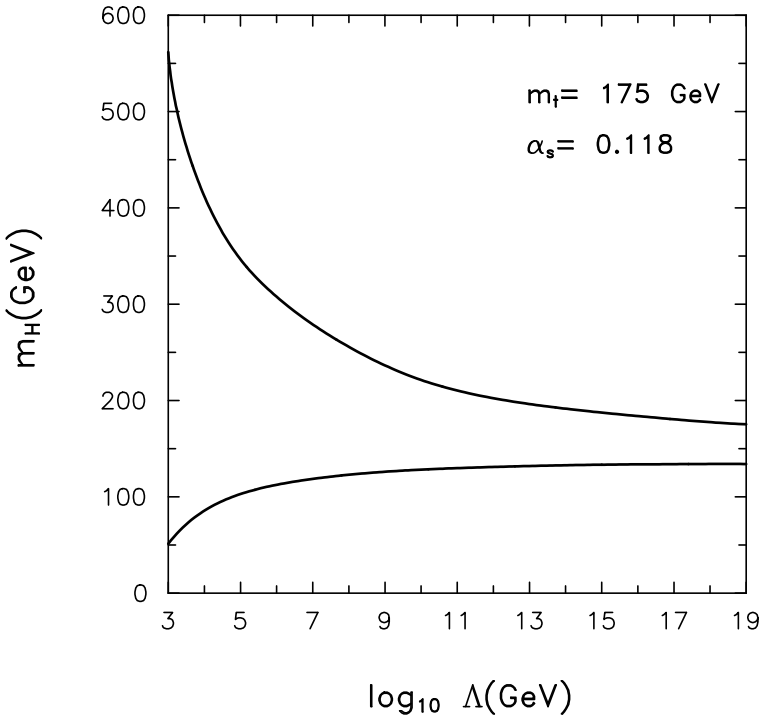


Figure 2.7: Bounds on the standard-model Higgs-boson mass that follow from requirements that the electroweak theory be consistent up to the energy Λ. The upper bound follows from triviality conditions; the lower bound follows from the stability conditions. Also shown is the range of masses permitted at the 95% confidence level by precision measurements and direct searches [11].

Figure 2.7 shows the triviality bound (upper curve) and the stability bound (lower curve) on the Higgs boson mass as a function of the energy scale Λ. The upper bound of Higgs mass is provided by the analysis of the triviality of the Higgs potential [20]. The running coupling constants λ has a functional dependence on the energy scale Λ from the renormalization group equation. If m_H is large, the upper bound manifests itself in the (one-loop) Landau pole, the value of the energy scale where the

theory becomes non-perturbative. The stability bound is obtained by computing quantum corrections to the classical Higgs potential and requiring that $< \phi_0 >= v/\sqrt{2}$ be an absolute minimum of the Higgs potential [21]. It is equivalent to requiring the quartic self-coupling constant of the Higgs potential $\lambda > 0$.

If we expect that the Standard Model is valid up to a given scale - let us say $\Lambda_{\mathrm{GUT}} \sim 10^{16}$ GeV - a bound on the Higgs mass should lie between both curves, in this case 140 GeV/c^2 $\lesssim M_H \lesssim$ 180 GeV/c^2.

Experimental Constraint

At LEP-2 the Higgs boson was primarily searched for in the production channel in which it is radiated off a virtual Z boson. The fact that no clear signal for Higgs boson production has been observed at LEP-2 has lead to a lower limit of the possible mass of the Higgs boson of $m_{\mathrm{H}} > 114.4$ GeV/c^2 at 95% C.L. by combining the searches in all possible final state channels by all LEP experiments [22].

Indirect constraints on the Higgs mass come from electroweak measurements in the framework of the SM. The radiative corrections to the parameter ρ, the relative strength of the charged current and neutral current, give us the functional dependence of m_W on m_t and m_H [23]. The definition of ρ is written as:

$$\rho = \frac{m_W^2}{m_Z^2 \cos^2 \theta_W} = \frac{1}{1 - \delta\rho}. \tag{2.21}$$

The one loop correction due to a $t - b$ loop in diagram 2.8(a) is:

$$\delta\rho = \frac{3G_F}{8\sqrt{2}\pi^2}[m_t^2 + m_b^2 - \frac{2m_t^2 m_b^2}{m_t^2 - m_b^2} \ln \frac{m_t^2}{m_b^2}]. \tag{2.22}$$

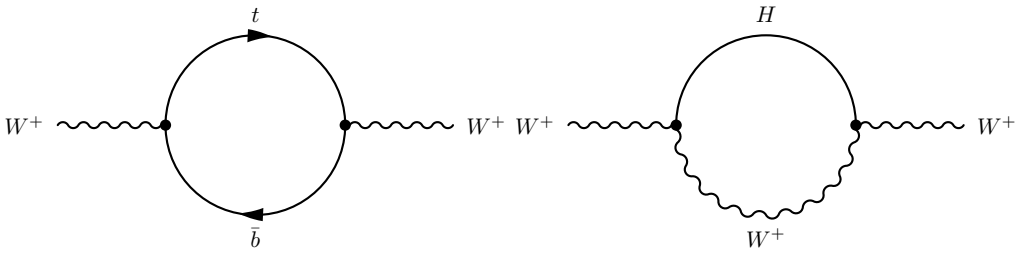


Figure 2.8: Feynman diagram of one loop correction of W mass due to (a) $t - b$ loop, and (b) Higgs loop.

The one loop correction due to a $H - W$ loop depicted in Diagram 2.8(b) in the $m_H \gg m_W, m_Z$ limit is:

$$\delta\rho = -\frac{3G_F}{8\sqrt{2}\pi^2}[\frac{m_w^2 \sin^2\theta_W}{\cos^2\theta_W} \ln\frac{m_H^2}{m_W^2}]. \tag{2.23}$$

It shows the W boson mass correction is proportional to m_t^2 and $\ln(m_H^2)$. Figure 2.9(a) compares the direct measurements of m_W and m_t (LEP2 and Tevatron data) with the indirect determination through electroweak radiative corrections (LEP1 and SLD) [24]. It shows that the electroweak measurements prefer a low mass Higgs. The uncertainty of the direct top mass measurement is $\sim$ 3 GeV/c^2 which is crucial to constraining the Higgs mass as shown in Figure 2.9(b). Taking all direct and indirect data into account, one obtains the best constraints on m_H. The global electroweak fit result in the $\Delta\chi^2 = \chi^2 - \chi^2_{\rm min}$ curve is shown in Figure 2.10. The χ^2 per degree of freedom is 18.2/13. Combining with the lower limit on m_H, the indirect Higgs mass constraint at 95% C.L. is

$$114.4 \text{ GeV} \ < \ m_H \ < \ 144 \text{ GeV}. \tag{2.24}$$

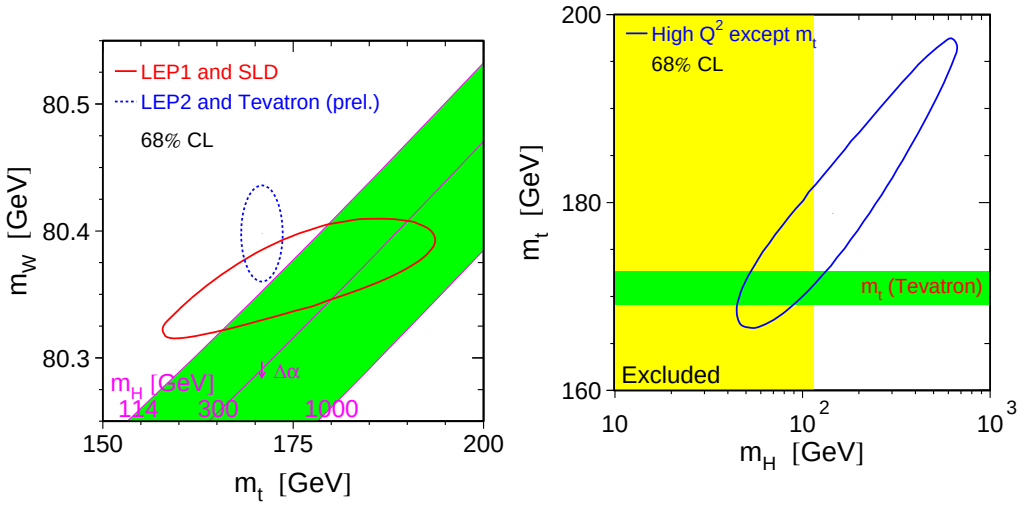


Figure 2.9: (a) Comparison of the direct measurements of m_W and m_t (LEP2 and Tevatron data) with the indirect determination through electroweak radiative corrections (LEP1 and SLD). Also shown in the SM relationship for the masses as function of m_H. (b) makes the analogous comparison for m_t and m_H.

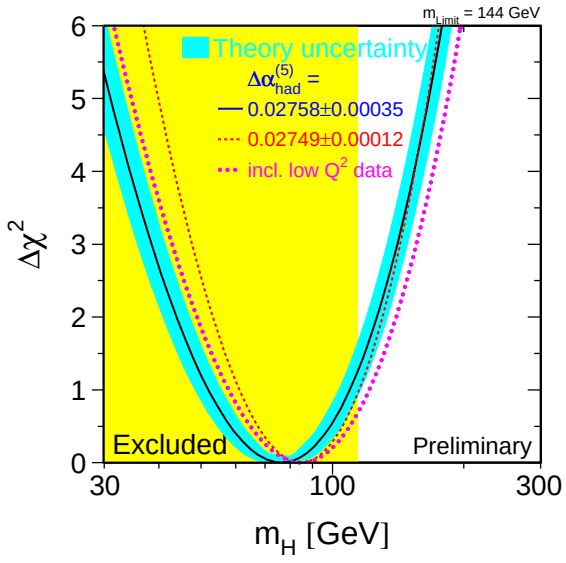


Figure 2.10: The indirect constraint of m_t vs m_H compares to the direct top mass measurement.

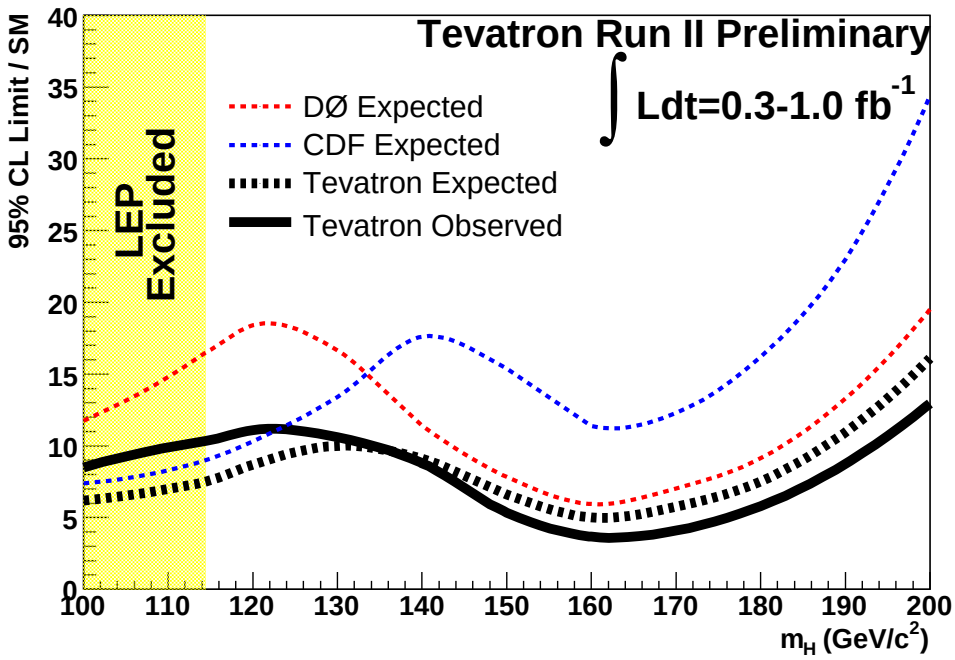


Figure 2.11: CDF and DØ's observed 95% CL limits on the Higgs production cross-section, divided by the corresponding SM predictions, for each of the five search channels in 2006 [25].

The current best place to look for a Higgs boson in the mass range from 114 to 144 GeV/c^2 is at the Tevatron. However, the production cross-section for Higgs at the Tevatron is very small, 1 pb to 10 fb for m_H between 110 $\sim$ 200 GeV/c^2, while the background in a hadron collider such as the Tevatron can be a factor of O(2) or O(9). The observed 95% CL limits in all of CDF's SM Higgs channels in the end of 2006 are shown in Figure 2.11 [4], as well as the limits from DØ's channels [26], The result is presented as the ratio of observed cross-section limits to the SM prediction. The direct Higgs search at CDF is about factor of 12 larger than SM predicted production cross-section for $m_H = 160$ GeV/c^2. The thesis discussed here leads to a factor of 5 improvement over this limit.

2.2.3 Spin-0 Physics

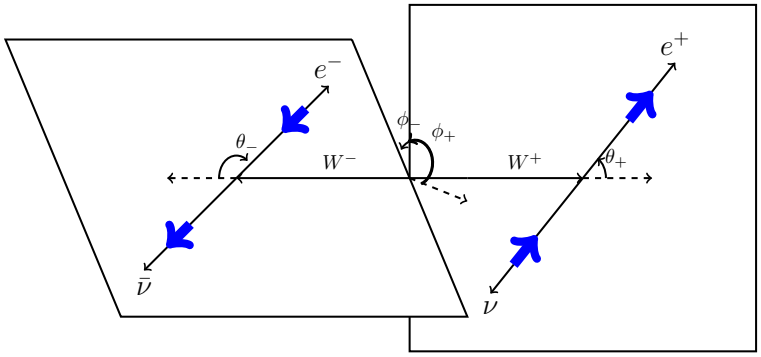


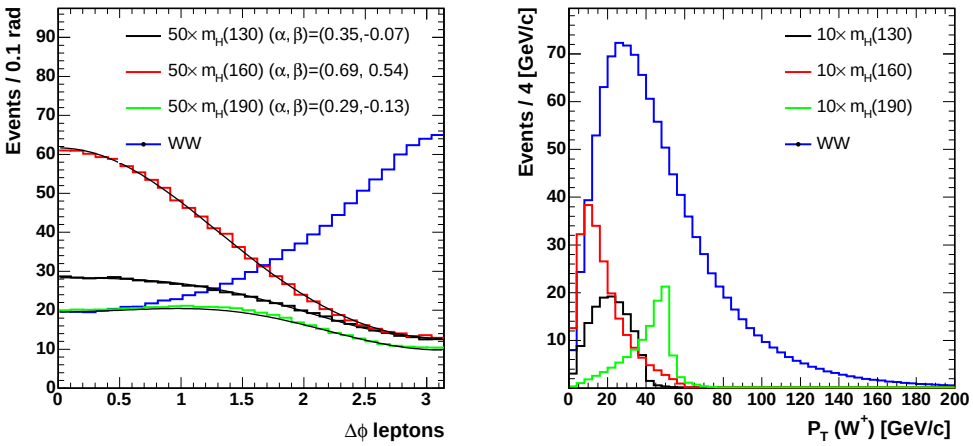
Figure 2.12: The decay plane of $H \to WW^*$ when both Ws decay leptonically.

Because the Higgs boson is a spin zero particle, there are interesting kinematic differences that can be expoited to discriminate it from backgrounds. Figure 2.12 shows the decay daughters of W in the Higgs center-of-mass frame by choosing a decay axis. The polar angle $\theta_+(\theta_-)$ depicted here is the angle of lepton in W rest frame. $\phi_+(\phi_-)$ is the azimuthal angle of the $W^+(W^-)$ decay plane measured in the H frame. The V-A structure of the charged current in weak decays produces correlations between the final state leptons and the W polarization. The polar angular distributions of $e^\pm$ for different $W^\pm$ helicity states are:

$$\begin{array}{llll} W^+(\lambda_+ = 1) & : (1+\cos\theta_+)e^{i\phi_+} & W^-(\lambda_- = 1) & : (1-\cos\theta_-)e^{-i\phi_-} \\ W^+(\lambda_+ = 0) & : \sin\theta_+ & W^-(\lambda_- = 0) & : \sin\theta_- \\ W^+(\lambda_+ = -1) & : (1-\cos\theta_+)e^{-i\phi_+} & W^-(\lambda_- = -1) & : (1+\cos\theta_-)e^{i\phi_-} \end{array} \quad (2.25)$$

These equations describe the feature that the positron tends to decay along the W^+ positive helicity direction while the electron tends to move along the negative helicity direction of W^-. Positive helicity is defined as the spin direction pointing along the direction of motion.

Because of the spin zero nature of scalar Higgs, the decayed W^+W^- pairs have to be anti-correlated in spin projection and end up with three helicity combinations $(\lambda_+, \lambda_-) = (+1, +1), (0, 0), (-1, -1)$. In the case of (1,1), the spin direction of $W^+(W^-)$ tend to point to right(left) in Figure 2.12. The decay daughter $e^+(e^-)$ tend to both distribute to right direction in each Ws rest frame. e^+ is then boosted further forward along the W^+ moving direction and the p_T in lab frame will be harder. e^- is boosted to the left such that the p_T in lab frame will be softer. The opening angle of two leptons still tend to be collinear. In the case of (-1,-1), the spin direction of $W^+(W^-)$ tend to point to left(right) which against their moving direction. The decayed leptons $e^+(e^-)$ tend to both decay to left in each Ws center-of-mass frame. The boosted P_T in lab frame for $e^+(e^-)$ is softer(harder). In the (0,0) state, where both W^+W^- are longitudinal polarized, two leptons both distribute as $\sin\theta$ distribution in each W's reset frame. It ends up with no specific small open angle of two leptons in this polarization state.



(a) (b)

Figure 2.13: The azimuthal angular distribution of two leptons and p_T distributions of W in the WW center of mass frame.

The azimuthal angle of the two leptons distributs according to the functional

form of

$$1 + \alpha\cos(\phi) + \beta\cos(2\phi). \tag{2.26}$$

The exact calculations of coefficient α and β are shown in the literature [25]. Figure 2.13(a) shows the $\Delta\Phi(l^+l^-)$ distributions of three different Higgs masses and standard model background $qq \to WW$. The parameter α and β are fitted and shown in the plot. The fitted values deviate from the calculations in [25], which assumes both W are on-shell but our Monte Carlo has zero-width approximation and smear the P_T of W a bit. Figure 2.13(b) shows that the larger the Higgs mass difference to $2m_W$, $|m_H - 2m_W|$, the harder the P_T spectrum of W is. Note that there is a upper cutoff of W P_T for $m_H < 2m_W$. The extra W P_T smears the open angle of two leptons decaying from each W. It is $m_H = 160$ GeV/c^2 gives us the smallest opening angle distribution and the softest W p_T spectrum.

The continuum $q\bar{q}' \to WW$ production also has constraints due to the spin correlation of the two initial quarks. The z component of the two quarks could be +1, -1 and 0. In the case of 0 spin components, the continuum WW will behave as the scalar Higgs boson which has one longitudinal W pair $W_L^+W_L^-$ or two transverse W pairs $W_T^+W_T^-$. However, the +1 and -1 gives the extra W pairs $W_L^+W_T^-$ or $W_T^+W_L^-$. After spin average on initial quarks, the final dilepton azimuthal angle distribution is different from Higgs boson as shown in the Figure 2.13. It is a good property to be used for Higgs search.

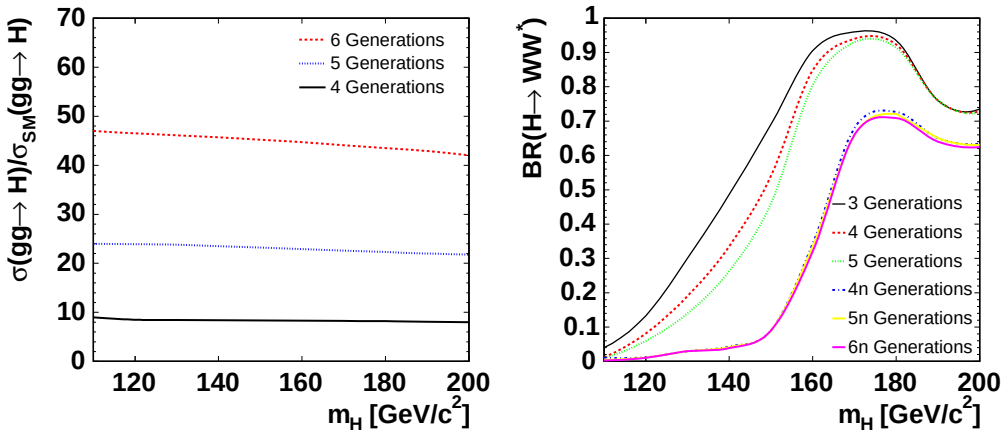
2.2.4 New Physics Models

The production cross-section of $H \to WW$ could be enhanced if new Physics beyond SM exists. The enhancement rate could be as large as a factor 100 of SM predictions. Tevatron has sensitivity to probe new Physics before testing SM. Here we will briefly describe three different models. The comparison to experimental results are shown in Chapter 6.4.1.

Extra Generations

The number of fermion generations is not fixed by the SM [27]. Quarks from the extra generations contribute to the loop mediated process in the Higgs boson production $gg \to H$ at the hadron colliders. The resulting enhancement of cross-section is a factor

of $|I_t + \sum I_{u_n} + I_{d_n}|^2/|I_t|^2$, where I_t, I_{u_n} and I_{d_n} are the quark amplitudes of top, n_{th} generation of up-type quark and down-type quark. In the limit of infinitly heavy quarks, the expected enhancement factors for 4, 5, 6 and n generation are 9, 25, 49 and $|1 + 2 \times (n-3)|^2$. Figure 2.14(a) shows the production cross-section of various extra families. The decay width of all lop mediated processes, $H \to \gamma\gamma$, $H \to Z\gamma$ and $H \to gg$, are altered and affect the branching ratio of $H \to WW^*$ as shown in Figure 2.14(b). If we have extra generation of heavy neutrino N, we also have to consider the effect of new channel, namely $H \to N\bar{N}$. Details on how the branching ratio of all Higgs decay channels change for extra SM generations can be found in [28]. In Figure 2.14(b), 4n, 5n and 6n denote the cases of one, two and three extra SM generations with neutrinos of masses ≈ 50 GeV, respectively.



(a) (b)

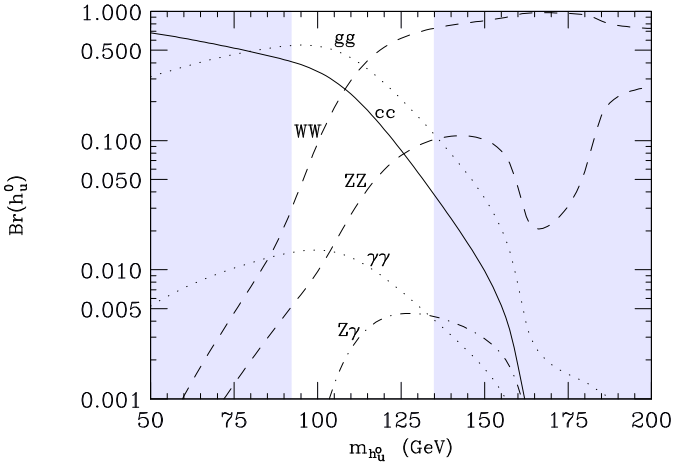
Figure 2.14: (a) production cross-section ratio to SM. (b) branching ratio of $H \to WW^*$ in various extra generation models [27]

Two Higgs Doublet

In the two Higgs doublet model, three Goldstone bosons are eaten and give longitudinal degree of freedom to three weak bosons. There are five physical Higgs-like particles after electroweak symmetry breaking. In the type II Higgs doublet model, the two CP even Higgs, h_u and h_d couple to up-type and down-type family separately. The ratio of the vacuum expectation values is

$$\tan\beta = \frac{h_u}{h_d}. \tag{2.27}$$

We can choose large $\tan\beta$ limits such that the physical Higgs doesn't mix after Spontaneous Symmetry Breaking. The lighter CP even Higgs equals to h_u [29]. h_u mass is always light (between m_Z and 135 GeV/c). By assumption it cannot decay into down-type family, $b\bar{b}$ nor $\tau^+\tau^-$. It is only allowed to decay into $c\bar{c}$, gg, $\gamma\gamma$, WW^*, $Z\gamma$ and ZZ^*. h_u acts like the standard model Higgs in every way except it does not couple to down type quarks nor leptons. The decay branching fractions for this state are shown in Figure 2.15(a). The shaded region is not allowed in the high $\tan\beta$ limit.



(a)

Figure 2.15: Branching ratio of h_u [29], up-quark coupling only Higgs in two Higgs doublet model. The shaded region is not allowed in the high $\tan\beta$ limit.

Manohar-Wise Model

In [30] Manohar and Wise argue that dimension 6 operators may contribute significantly to Higgs production and decay processes for which loop processes dominate in the standard model. E.g., they calculate explicitly the ratio of cross-sections for $gg \to H$ with and without those contributions as a function of the Wilson coefficient of the operator in the effective theory. Figure 2.16 is the result of adding a dimension six operator into the SM.

$$\delta L = -\frac{c_G g_s^2}{2\Lambda^2} H^\dagger H G^a_{\mu\nu} G^{a\mu\nu}, \tag{2.28}$$

where $G_{\mu\nu}$ is the gluon field strength tensor and H is the Higgs doublet. Λ is the energy scale of new physics. Expanding about the Higgs field with vacuum expectation value

$v \approx 250$ GeV and only keep the term linear in h which contributes to the production rate for Higgs bosons.

$$\delta L = -\frac{c_G g_s^2 v}{2\Lambda^2} h G^a_{\mu\nu} G^{a\mu\nu}. \tag{2.29}$$

The operator leads to additional vertices and modify the $h \to gg$ decay amplitude. The new physics effects can modify factor of 2 or more for the $gg \to h$ production rate and the $h \to gg$, $h \to \gamma\gamma$ and $h \to \gamma Z$ decay rate all of which have one loop SM amplitudes.

They find a factor 2-3 or more enhancement of the cross-section roughly independent of Higgs mass, and without any significant change to the $H \to WW$ branching fraction. This enhancement is not present for associate production, e.g. WH and ZH, as those proceed at tree level in the standard model. For CDF this means that only $H \to WW$ search benefits from this type of new physics as all other final states, most notably $b\bar{b}$, require associate production to be visible on top of background! They conclude their discussion by noting that one might expect the size of the coefficient to be O(0.1) based on gauge coupling unification in GUT theories at O(10^{15}) GeV as motivated by neutrino masses.

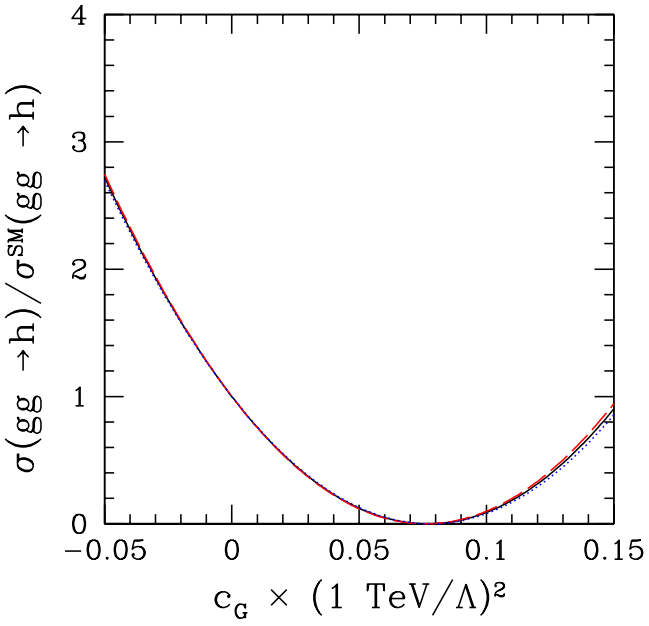


Figure 2.16: The ratio of the $gg \to H$ cross-section to its SM value as a function of C_G for $m_H = 120$ GeV/c^2 (dashed red), $m_H = 140$ GeV/c^2 (solid black) and $m_H = 160$ GeV/c^2 (dotted blue) [30].

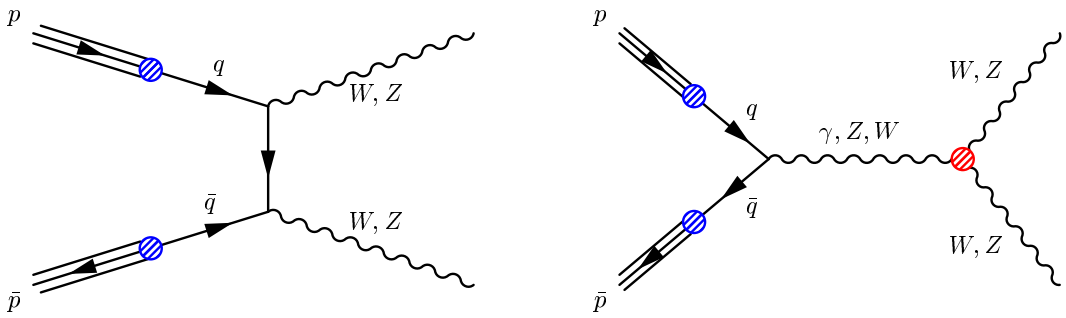


Figure 2.17: Dominant diagrams in the production of boson pairs: (a) t-channel (b) s-channel.

2.3 Diboson Physics: Trilinear Gauge Boson Couplings

The SM Lagrangian, Equation (2.4), specifies the gauge boson self-interactions. We can re-write $W^{1,2,3}$ and B field into physical field $W^{\pm}, Z$ and A.

$$\begin{aligned}
\text{First two terms} &= -\tfrac{1}{4}W^a_{\mu\nu}W_a^{\mu\nu} - \tfrac{1}{4}B_{\mu\nu}B^{\mu\nu} \\
&= -\tfrac{1}{16}(W^+ + W^-)_{\mu\nu}(W^+ + W^-)^{\mu\nu} \\
&\quad +\tfrac{1}{16}(W^+ - iW^-)_{\mu\nu}(W^+ - iW^-)^{\mu\nu} \\
&\quad -\tfrac{1}{4(g_1^2+g_2^2)}(g_2 Z + g_1 A)_{\mu\nu}(g_2 Z + g_1 A)^{\mu\nu} \\
&\quad -\tfrac{1}{4(g_1^2+g_2^2)}(-g_1 Z + g_2 A)_{\mu\nu}(-g_1 Z + g_2 A)^{\mu\nu}
\end{aligned} \tag{2.30}$$

From this equation, we can observe that there is no triple gauge couplings between neutral gauge bosons. There are only quartic coupling terms in the form of $WWWW$, $WWAA$, $WWZZ$ or $WWAZ$. The charged triple gauge coupling can be derived by picking terms WWZ or $WW\gamma$. The measurement of triple gauge boson vertices is a direct test of the gauge structure of the $\mathrm{SU(2)_L \times U(1)_Y}$ theory.

By comparing the data with a general WWV vertex with $V = \gamma, Z$, there are possibilities to observe anomalous gauge boson couplings that can be induced, for instance, by radiative corrections in the SM or by New Physics effects. The dominant contributions to the cross-sections for WW and WZ production are the t-channel (and similar u-channel) process involving two instances of the well measured boson-quark couplings and the s-channel process involving triple gauge couplings, shown in Figure 2.17. These vertices have been tested at LEP2 with percent level precision up to the $2m_W$ energy scale. A significant feature of the diboson production at the Tevatron is at energies far above the LEP energies. Trilinear Gauge coupling measurements at the Tevatron thus probes different physics than LEP.

The most generalized charged triple gauge boson coupling Lagrangian [31] can be written as

$$\begin{aligned}
\tfrac{1}{ig_{WWV}}\mathcal{L}_{\text{eff}}^{WWV} &= g_1^V V^\mu \left(W^-_{\mu\nu}W^{+\nu} - W^+_{\mu\nu}W^{-\nu}\right) \\
&+ \kappa_V\, W^+_\mu W^-_\nu V^{\mu\nu} \\
&+ \tfrac{\lambda_V}{M_W^2}\, V^{\mu\nu} W_\nu^{+\rho} W^-_{\rho\mu} \\
&+ ig_5^V \varepsilon_{\mu\nu\rho\sigma}\left((\partial^\rho W^{-\mu})W^{+\nu} - W^{-\mu}(\partial^\rho W^{+\nu})\right) V^\sigma \\
&+ ig_4^V W^-_\mu W^+_\nu (\partial^\mu V^\nu + \partial^\nu V^\mu) \\
&- \tfrac{\tilde\kappa_V}{2} W^-_\mu W^+_\nu \varepsilon^{\mu\nu\rho\sigma} V_{\rho\sigma} \\
&- \tfrac{\tilde\lambda_V}{2m_W^2}\, W^-_{\rho\mu} W^{+\mu}{}_\nu \varepsilon^{\nu\rho\alpha\beta} V_{\alpha\beta}
\end{aligned} \tag{2.31}$$

with the overall couplings defined by $g_{WW\gamma} = e$ and $g_{WWZ} = e\cot\theta_W$ and where the reduced field strengths $W_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu$ and $V_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu$ are used. For on–shell photons, i.e. with $q^2 = 0$, the couplings $g_1^\gamma = 1$ and $g_5^\gamma = 0$ are fixed by $\mathrm{U(1)_Q}$ gauge invariance. In the Lagrangian eq. (2.31), the couplings g_1^V, κ_V and λ_V separately conserve C and P symmetries, while g_5^V violates them but conserves CP symmetry. The couplings g_4^V, $\tilde\kappa_V$ and $\tilde\lambda_V$ parameterize a possible CP violation in the bosonic sector. Note that the C and P conserving terms in $\mathcal{L}_{\text{eff}}^{WW\gamma}$ correspond to the lowest order terms in a multiple expansion of the W boson–photon interactions, the charge Q_W, the magnetic dipole moment μ_W and the electric quadrupole moment q_W of the W^+ boson [31]

$$Q_W = eg_1^\gamma \ , \quad \mu_W = \frac{e}{2M_W}\left(g_1^\gamma + \kappa_\gamma + \lambda_\gamma\right) \ , \quad q_W = -\frac{e}{M_W^2}\left(\kappa_\gamma - \lambda_\gamma\right) \tag{2.32}$$

In the SM, at the tree level, the trilinear couplings are simply given by $g_1^Z = g_1^\gamma = \kappa_Z = \kappa_\gamma = 1$, while all the other couplings in eq. (2.31) are zero. It became common practice to introduce deviations of the former set of couplings from their tree-level SM values

$$\Delta g_1^Z \equiv (g_1^Z - 1) \ , \quad \Delta\kappa_\gamma \equiv (\kappa_\gamma - 1) \ , \quad \Delta\kappa_Z \equiv (\kappa_Z - 1) \tag{2.33}$$

The aTGC limits at LEP has been constrained to 3% level. However, aTGC measurements at hadron colliders are probing higher center of mass energy regime. The aTGC results are reported in terms of these three parameters at different energy scale Λ. To avoid unitarity violation in the WWZ vertex, a dipole form factor is introduced as a cut-off by multiplying the coupling with $1/(1 + s/\Lambda^2)^2$ where s is the center-of-mass energy squared. Assuming that only one anomalous coupling is nonzero at a time, one

finds, for $\Lambda \gg m_W, m_Z$ [32]

$$\begin{array}{ll} |\Delta g_1^Z| \leq \frac{0.84\text{TeV}^2}{\Lambda^2} & \\ |\Delta\kappa_\gamma| \leq \frac{1.81\text{TeV}^2}{\Lambda^2} & |\Delta\kappa_Z| \leq \frac{0.83\text{TeV}^2}{\Lambda^2} \\ |\lambda_\gamma| \leq \frac{0.96\text{TeV}^2}{\Lambda^2} & |\lambda_Z| \leq \frac{0.96\text{TeV}^2}{\Lambda^2}. \end{array} \tag{2.34}$$

The bounds listed in Equation 2.34 is computed with $m_W = 80$ GeV/c^2, $m_Z = 91.1$ GeV/c^2. If more than one coupling is varied at a time, cancellations between the TGC's may occur, and the unitarity limits are weakened. Experiments constrain the TGC couplings non-trivially only if the experimental limits are more stringent than the unitarity bounds, for a given value of Λ.

2.4 Hadron Collider Physics

The theoretical calculations of hadron collider processes are complicated due to the internal structure of colliding hadrons. The parton model is proposed to factorize the calculation into perturbative parts (hard scattering), and non-perturbative parts (parton momentum distribution). This successfully describes the collisions in hadron colliders. Other complexities of hadron collisions, such as multiple interactions per beam crossing and spread from parton fragments, will also be mentioned in this section. A detailed review can be found in the literature [33].

2.4.1 Parton Model

The parton model was proposed by Richard Feynman in 1969 as a way to analyze high-energy hadron collisions. In this model, a hadron (for example, a proton) is composed of a number of point-like constituents, termed "partons". Additionally, the hadron is in a reference frame where it has infinite momentum, a valid approximation at high energies. Thus parton motion is slowed by time dilation, and the hadron charge distribution is Lorentz-contracted, so incoming particles will be scattered instantaneously and incoherently. The QCD factorization theorem states that the cross-sections for high energy hadronic reactions with a large momentum transfer can be factorized into a parton-level "hard scattering" convoluted with the parton "distribution functions". For scattering of two hadrons A and B to produce a final state X of our interest, the cross-section can be formally written as a sum over the sub-process cross-sections from the

contributing partons

$$\sigma_{AB} = \sum_{a,b} \int dx_a dx_b \; f_{a/A}(x_1, Q^2) f_{b/B}(x_b, Q^2) \hat{\sigma}(ab \to X), \tag{2.35}$$

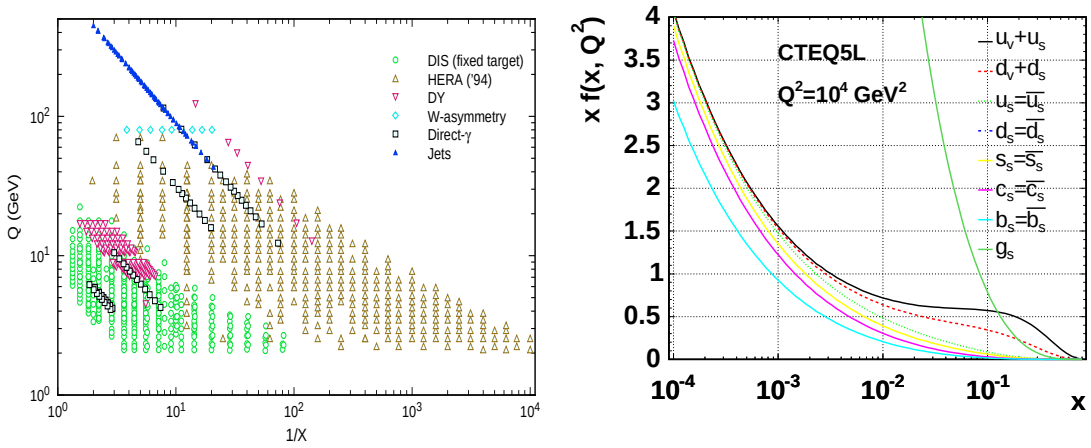
where Q^2 is a large momentum scale that characterizes the hard scattering, e.g. $m^2_{l^+l^-}$ and p_T^2. The typical Q^2 in the hard scattering process is much larger than QCD energy scale $\Lambda^2_{QCD} \approx (200 \text{ MeV/c})^2$. The parton-level hard scattering cross-section can be calculated perturbatively in QCD to $O(\alpha_S^n)$, while the parton distribution functions parameterize the non-perturbative aspect. Equation (2.35) can be written as

$$\sigma_{AB} = \sum_{a,b} \int dx_a dx_b \; f_{a/A}(x_1, \mu_F^2) f_{b/B}(x_b, \mu_F^2) [\hat{\sigma}_0 + \alpha_S(\mu_R^2)\hat{\sigma}_1 + ...]_{ab\to X}. \tag{2.36}$$

Here μ_F is the factorization scale which separates hard and soft physics and μ_R is the renormalization scale for the QCD running coupling. The cross section is invariant under change parameters μ_F and μ_R since higher order cross section term $\hat{\sigma}_n$ can compensate the scale dependence of parton density function and running coupling α_S. Normally, the values of μ_F and μ_R are chosen to be the order of the typical momentum scales of the hard scattering processes. For example, the standard choice $\mu_F = \mu_R = m_{l^+l^-}$ is chosen for Drell-Yan process ($f\bar{f} \to Z/\gamma^* \to l^+l^-$). A next-to-leading order calculation is calculated up to $O(\alpha_S)$.

Parton density functions have x and Q^2 dependence and are measured by global fitting from $p\bar{p}$, ep or γp experiments. Figure 2.18(a) shows the kinematics region covered by data sets in CTEQ5 global fit [34]. The high x low Q regions are predominantly measured via deep inelastic scattering (DIS) or Drell-Yan processes in fixed target (ep and γp) experiments. The low x regions are mostly measured in HERA (Hadron Electron Ring Accelerator)with 27.5 GeV electrons or positrons colliding on 920 GeV protons [35]. The high x gluon distribution is mainly coming from analyzing jet transverse energy distribution in the $p\bar{p}$ collider experiments, CDF and D0. Figure 2.18(b) shows Parton Density Functions for various types of partons at momentum scales $Q = 100$ GeV.

The valence quarks u_v, d_v, as well as the gluons carry a large momentum fraction, typically $x \sim 0.08 - 0.3$. The "sea quarks" ($\bar{u} = u_{sea}, \bar{d} = d_{sea}, s, c, b$) have small x, and are significantly enhanced at higher Q^2. Both of these features lead to



(a) (b)

Figure 2.18: (a) Kinematic map of the (x, Q) range covered by the data sets used in CTEQ5 global analysis [34]. (b) Parton density functions versus parton energy fraction x at momentum scale $Q = 100$ GeV, from CTEQ-5L.

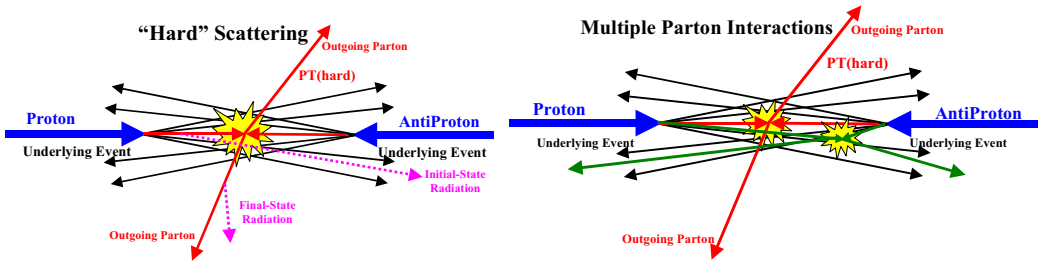
important collider consequences. Consider the center-of-mass energy of given particles E_{cm},

$$E_{cm}^2 = x_1 x_2 \hat{s}, \tag{2.37}$$

where x_1, x_2 are momentum fractions of partons in two beams and $\hat{s}$ is the center-of-mass energy squared of beam. Heavy objects near the beam energy threshold have high x and are more likely produced via valence quarks. For a given mass of heavy objects, the higher the beam center-of-mass energy ($\hat{s}$), the lower x it is. In low x, the sea quark PDFs are dominant over valance quarks.

2.4.2 Pile-up Effect and Underlying Event

The pileup effect can be categorized into two main effects [36]. One coming from the noise and "out-of-time" signal from electronics, called zero-bias event. The other comes from secondary interactions occurring during the beam crossing, called minimum bias events. Both are linearly dependent on the instantaneous luminosity. A correct model of minimum bias event is important especially for high luminosity collisions. We can calculate the average number of minimum bias interactions according to the chosen luminosity setting and the time between bunch crossings. The number of minbias event is picked according to a Poisson distribution.



(a) (b)

Figure 2.19: (a) shows a 2-to-2 parton-parton scattering with transverse momentum, p_T(hard). (b) illustrates the way PYTHIA models the underlying event in proton-antiproton collision by including multiple parton interactions. There is a second "semi-hard" 2-to-2 parton-parton scattering that contributes particles to the underlying event.

Underlying event is defined as everything except the "hard scattering" at each collision. To describe how underlying event is simulated, we should understand how a single collision is simulated in PYTHIA [37]. Figure 2.19(a) illustrates the way QCD Monte-Carlo models simulate a proton-antiproton collision in which a "hard" 2-to-2 parton scattering with transverse momentum p_T(hard)has occurred. The resulting event consists of "hard scattering", th outgoing partons plus initial and final-state radiation, and beam-beam remnants, which come from the breakup of the proton and antiproton. The beam-beam remnants produce soft particles and correlates to the hard interactions.

FIG. 2.19(b) shows the way PYTHIA [37] models the underlying event in proton-antiproton collisions by including multiple parton interactions. In addition to the hard 2-to-2 parton-parton scattering and the beam-beam remnants, sometimes there is a second "semi-hard" 2-to-2 parton-parton scattering that contributes particles to the underlying event.

Chapter 3

Experimental environment

The Tevatron at Fermilab is the highest energy accelerator in the world during the data taking period for this thesis. Protons and anti-protons ($p\overline{p}$) each are accelerated to 980 GeV and collided with a center of mass energy of 1.96 TeV. The data used in this thesis is collected by the Collider Detector at Fermilab (CDF-II) at the collision point B0. There is another collider detector, D0, situated in the collision point D0. Both the Tevatron and its performance, as well as the CDF-II experiment will be discussed in this chapter.

3.1 Tevatron: $p\overline{p}$ collider

In order to create the world's most energetic particle beams, Fermilab uses a series of accelerators. Figure 3.1 shows the paths taken by protons and anti-protons from initial acceleration to collision in the Tevatron. The first stage of acceleration is in the Cockcroft-Walton pre-accelerator [38] , where hydrongen ions (H^-) are created by ionizing the hydrogen gas in the presence of a cesium surface and accelerated to a kinetic energy of 750 keV. The H^- ions enter a linear accelerator (Linac) [39], approximately 500 feet long, where they are accelerated to 400 MeV. The acceleration in the Linac is done by a series of "kicks" from Radio Frequency (RF) cavities. The oscillating electric field of the RF cavities groups the ions into bunches. Before entering the next stage, a carbon foil removes the electrons from the H^- ions at injection, leaving only the protons. The 400 MeV protons are then injected into the Booster, a 74.5 m-diameter circular synchrotron. The protons travel around the Booster about 20,000 times before reaching a final energy of 8 GeV. Protons are then extracted from the Booster into the Main

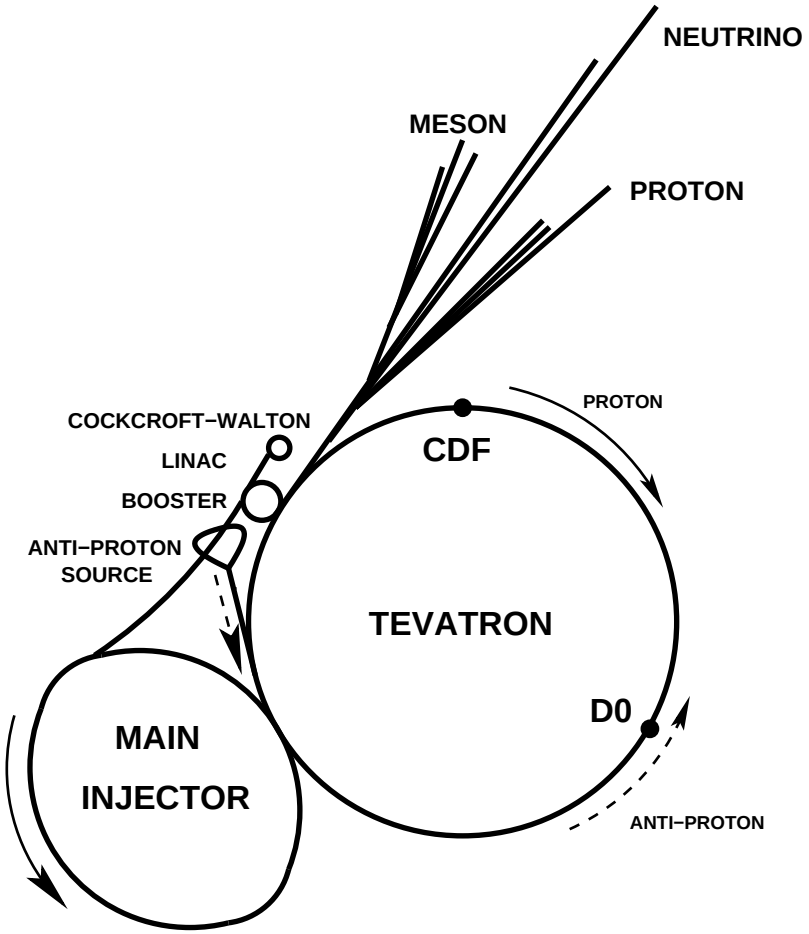


Figure 3.1: Layout of the Fermilab accelerator complex. The proton (solid arrow) is accelerated at the Cockcroft-Walton, Linac, Booster, Main Injector and finally at the Tevatron. The anti-proton (dashed arrow) from the anti-proton source is first accelerated at the Main Injector and then at the Tevatron.

Injector [40]. where they are accelerated from 8 GeV to 150 GeV.

Anti-protons are produced from proton beams in the Main Injector. The Main Injector also produces 120 GeV protons, colliding them with a nickel target to produce a wide spectrum of secondary particles, including anti-protons. In the collisions, about 20 anti-protons are produced per one million protons. The anti-protons are collected, focused, and then stored in the Accumulator ring. Once a sufficient number of anti-protons are produced, they are sent to the Main Injector and accelerated to 150 GeV. Finally, both the protons and anti-protons are injected into the Tevatron.

The Tevatron, the last stage of Fermilab's accelerator chain, receives 150 GeV protons and anti-protons from the Main Injector and accelerates them to 980 GeV. The protons and anti-protons travel around the Tevatron in clockwise and counterclockwise direction from top-view, respectively. The beams are brought to collision at the center of the two detectors, CDF-II and D0.

We use the term "luminosity" to quantify the beam particle density and the crossing rate, which is the number of particle collisions that occur each second. The luminosity, in units of $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, can be expressed as:

$$\mathcal{L} = \frac{f N_B N_p N_{\overline{p}}}{2\pi(\sigma_p^2 + \sigma_{\overline{p}}^2)} F\left(\frac{\sigma_l}{\beta^*}\right) \tag{3.1}$$

where f is the revolution frequency, N_B is the number of bunches, $N_{p/\overline{p}}$ are the number of protons/anti-protons per bunch, and $\sigma_{p/\overline{p}}$ are the RMS beam sizes at the interaction point. F is a form factor which corrects for the bunch shape and depends on the ratio of σ_l, the bunch length, to β^*, the beta function, at the interaction point. The beta function is a measure of the beam width, and is proportional to the beam's x and y extent in phase space. There are two ways to express the likelihood of collisions, or luminosity, during any particular "shot" of protons and anti-protons (store): peak luminosity, which describes the initial (and highest) luminosity in a store, and integrated luminosity, which describes how many total collisions are produced over the lifetime of that store. A store produces about 10 to 20 hours of collisions. Since an increase in total collisions provides more chances to produce new particles, high integrated luminosity is the key to discovery.

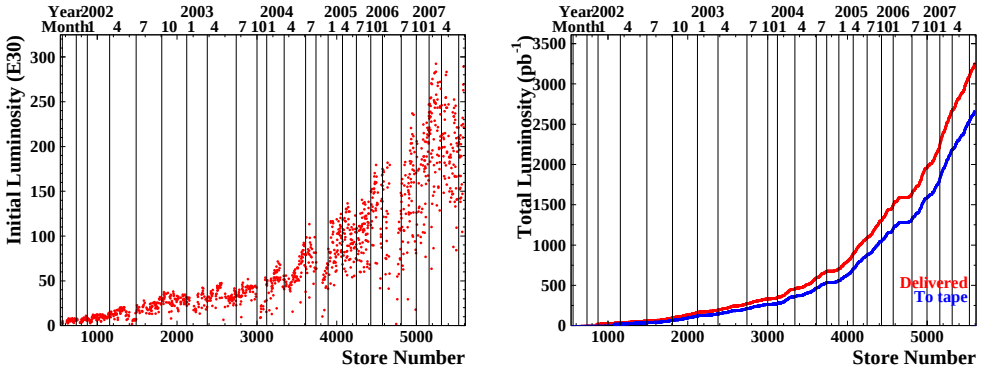
The Tevatron has reached design goal of the peak instantaneous luminosity, 2×10^{32} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, in the end of 2006. The maximum peak luminosity was recored to be $2.8503 \times 10^{32} cm^{-2} s^{-1}$ on 18 February 2007 with store number 5234. Figure 3.2(a) shows the peak luminosities for each store in CDF II and Figure 3.2(b) shows the delivered and recorded data in CDF II. Up to the writing of this thesis, there are 3.2 fb^{-1} delivered

Table 3.1: Characteristic parameters of the Tevatron and LHC.

Parameter	Tevatron	LHC
energy at center-of-mass	1.96 TeV	14 TeV
inter-bunch spacing	396 ns	25 ns
number of bunches N_B	36	2808
protons/bunch N_p	3×10^{11}	1.15×10^{11}
protons/bunch $N_{\bar{P}}$	3×10^{10}	-
bunch length σ_l	37 cm	7.55 cm
beam width β^*	35 cm	55 cm
interaction/crossing[a]	7	23
designed luminosity	$2 \times 10^{32} cm^{-2} s^{-1}$	1×10^{34}
peak luminosity	$2.8 \times 10^{32} cm^{-2} s^{-1}$	-
magnetic field	4.2 T	8.4 T
Super Conducting Dipoles	1000	1300
Ring Radius (km)	1	4.3

[a] At designed luminosity

by the Tevatron, and 2.7 fb^{-1} recorded in CDF II, while only 1.9 fb^{-1} are used for this analysis. The recorded efficiency is stable at 85% since 2003. The inefficiency is 5% each from three factors: trigger dead time and readout, start and end of each store and the problems from detectors or Data Acquisition (DAQ). Among the recorded data, there is additional 8% inefficiency due to the data quality requirement. The absolute efficiency of the data used in physics analysis is 78%.



(a) (b)

Figure 3.2: Accumulated luminosities for stores collided between April 2001 and August 2007. This analysis uses the data collected from March 23 2002 (store 1120) to March 30 2007(store 5314). The effective luminosity after the run quality cut is 1.9 fb^{-1}

3.2 The CDF-II Detector

The Collider Detector at Fermilab Run II (CDF-II) is a general purpose, azimuthally and forward-backward symmetric apparatus, designed to study $p\overline{p}$ collisions at the Tevatron. It can measure the properties of physics objects decayed from unstable particles in order to understand the hard interaction physics. The objects for interests are electrons, muons, quarks and neutrinos. Electrons and muons are charged particles which can be measured directly. Quarks exist in a form of a jet which is a fragmentation of multiple hadrons. Neutrinos carry away large momentum because of extremely low probability of interaction with detectors.

There are six subsystems of CDF-II most relevant to the dilepton+$\not\!\!E_T$ analysis: tracking, solenoid magnets, calorimetry, muon chambers, luminosity counter and trigger. Tracking and solenoid magnets can measure sign and momentum of charged particles. Calorimetry can measure deposit energies of particles. A large cluster of energies in calorimetry can be a jet and a large imbalance of energy is a measure of neutrino. Luminosity counter measures the number of $p\bar{p}$ interactions to normalize the rate of physics events. Trigger system is essential to identify and record the interesting events. This section will start from an overview of the CDF detector including the coordinate system, and described each sub-detectors and their performances.

3.2.1 Detector Overview

Because of its barrel-like detector shape, CDF-II uses a cylindrical coordinate system (r, ϕ, z) with the origin at the center of the detector. The z axis is along the direction of the proton beam. The r indicates the radial distance from the origin and ϕ is the azimuthal angle. The $r-\phi$ plane is called the transverse plane, as it is perpendicular to the beam line. The polar angle, θ, is the angle relative to the z axis. An alternative way of expressing θ, pseudo-rapidity (η), is defined as:

$$\eta \equiv -\ln\tan(\theta/2). \tag{3.2}$$

The pseudo-rapidity is a good unit since $\delta\eta$ is invariant under boosts along the z. The coverage of each CDF-II detector sub-system will be described using combinations of η, r, ϕ and z.

Figure 3.3 shows the detector and the different sub-systems in a solid cutaway view ($r - z$ plane). The innermost system of the detector is the integrated tracking

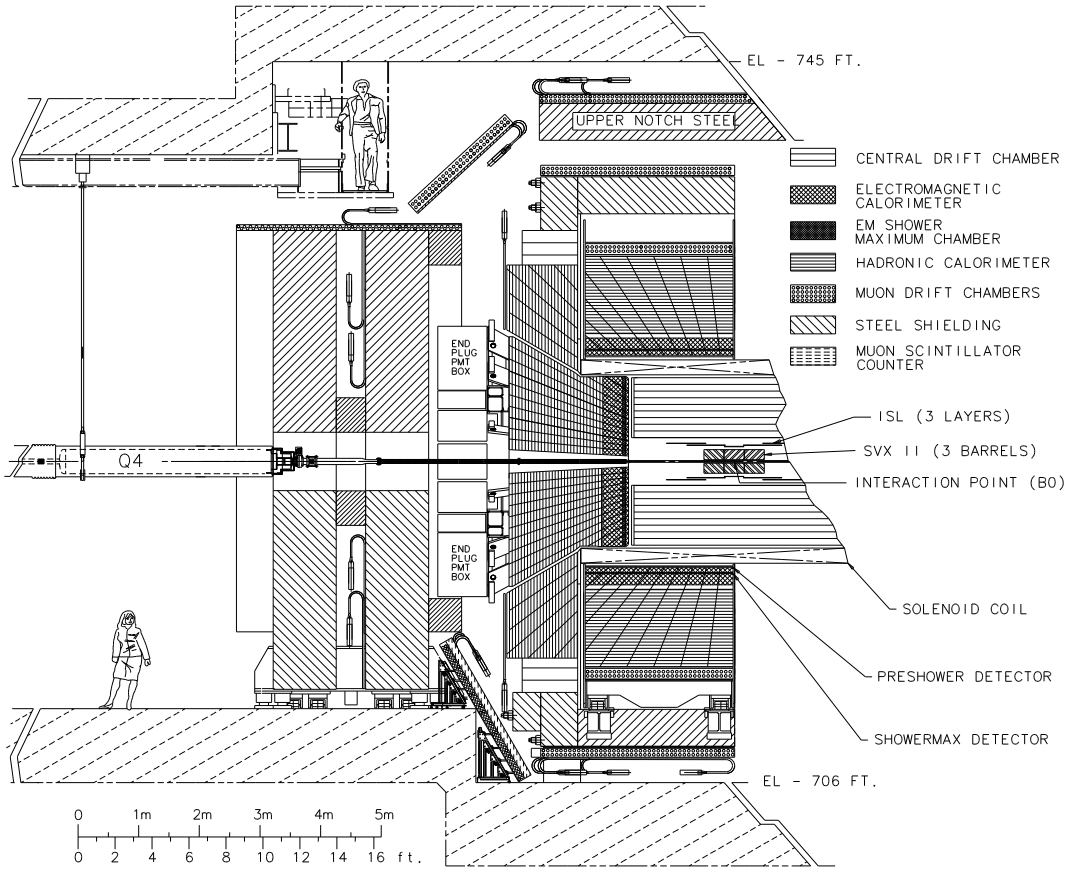


Figure 3.3: An elevation view of the CDF detector. In the central region we see the silicon vertex detector (SVXII) nested inside the open-cell wire drift chamber (COT). These tracking systems are contained in a 1.4 T solenoidal magnetic field. Then come electromagnetic and sampling calorimeters and muon systems, all of which are also visible in the high η region.

system: a silicon detector and a central outer tracker . The tracking system is designed to measure the trajectories of charged particles. Reconstructed particle trajectories are referred to as "tracks". Multiple-track reconstruction allows us to identify a vertex where either the $p\overline{p}$ interaction took place (primary vertex) or the decay of a long-lived particle took place (secondary or displaced vertex).

Immediately outside the COT is the Time of Flight system (TOF), which consists of 216 scintillator bars, roughly 300 cm in length and with a cross-section of 4×4 cm^2. The bars are arranged in a barrel around the COT outer cylinder. The TOF system is designed for identification of charged particles with momentum below 2 GeV/c. It is not used in this analysis since the minimum transverse momentum requirement is 10 GeV/c.

Both the tracking system and the TOF are contained in a superconducting solenoid, 1.5 m in radius, and 4.8 m in length, that generates a 1.4 Tesla magnetic field parallel to the beam axis (z). The field curves charged particle to be curved in the transverse plane (x-y) to measure the charge sign and momentum magnitude of tracks.

The solenoid is surrounded by the electromagnetic and hadronic calorimeters, which measure the energy of particles that shower when interacting with matter. The coverage of the calorimeters is $|\eta| < 3$. The electromagnetic calorimeter is a lead/scintillator sampling device and measures the energy of the electrons and photons. The hadronic calorimeter is an iron/scintillator device and measures the energy of the hadrons, e.g. pions and kaons.

The calorimeters are surrounded by the muon detector system. Muons interact with matter primarily through ionization. As a result, if a muon is created in the collision and has enough momentum, it will pass through the tracking system, TOF, the solenoid and the calorimeters with minimal interaction with the detector material. Muon detectors are, therefore, placed radially outside the calorimeters.

3.2.2 Tracking Systems

There are two types of tracking system in CDF-II - silicon detector and central outer tracker (COT). Silicon microstrip detectors is a semi-conductor detectors and record at most 7 hits for a charged track. COT is a drift chamber has 96 multi-layers. Figure 3.4 shows the $r - z$ view of the tracking volume and Figure 3.5 zoom in silicon detector.

Silicon Microstrip Detector

The silicon microstrip detector consists of three sub-detectors in a barrel geometry that extends from the radius of r= 1.35 cm to r= 28 cm and covers the track reconstruction in the range of $|\eta| < 2$. The $r - \phi$ view of the silicon tracker is shown in Figure 3.6. A single layer rad-hard Layer 00 detector is mounted on and supported by the beam pipe, with sensors at r=1.35 cm and r=1.62 cm. The next five concentric layers compose the silicon sensors (SVX II) and are double-sided detectors from r=2.45 cm to 10.6 cm. The axial side of each layer is used for r-ϕ measurements. The stereo side of each layer is used for r-z measurements. The two outer layers compose the Intermediate Silicon Layers (ISL) and are double-sided detectors, from r=20 cm to r=28 cm. This

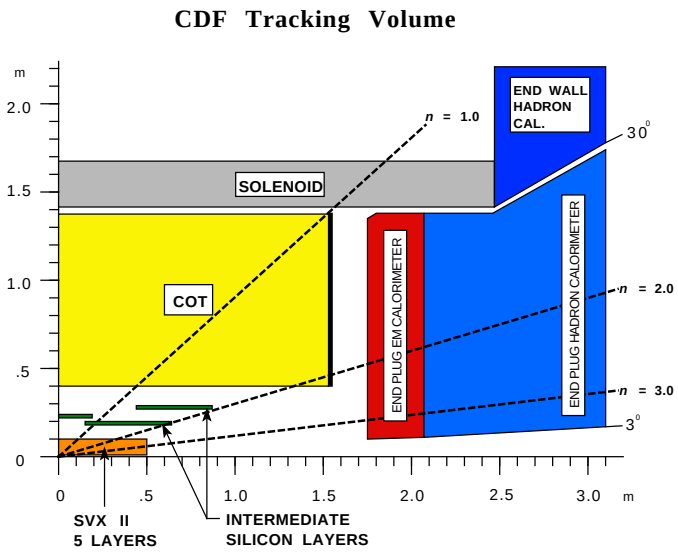


Figure 3.4: A diagram of the CDF-II tracker layout showing the tracking volume of tracking systems.

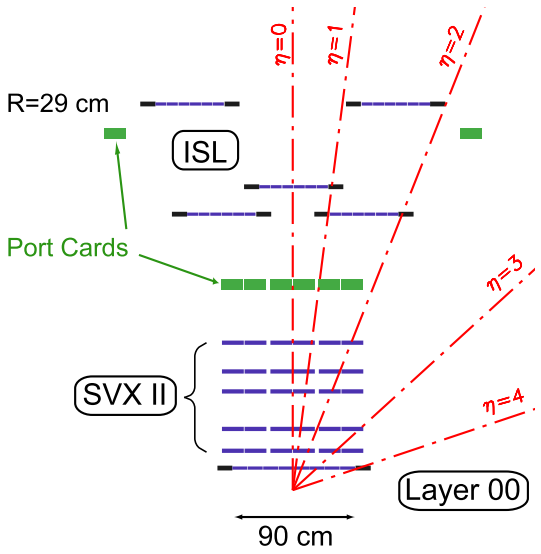


Figure 3.5: Coverage of the different silicon subdetector systems projected into the r-z plane. The r and z axes have different scales.

entire system allows charged particle track reconstruction in three dimensions. The impact parameter resolution of SVX II + ISL is 40 μm including 30 μm contribution from the beamline. The z_0 resolution of SVX II + ISL is 70 μm. The main parameters of the silicon tracker are summarized in Table 3.2.

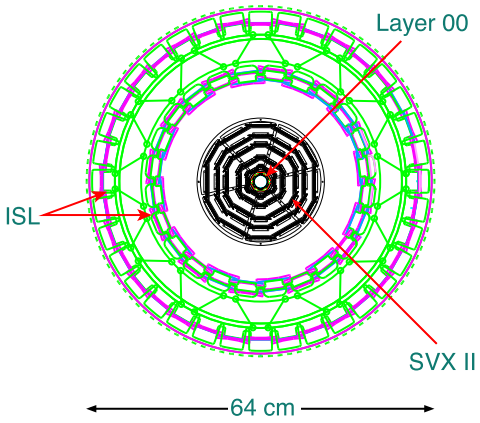


Figure 3.6: Silicon system.

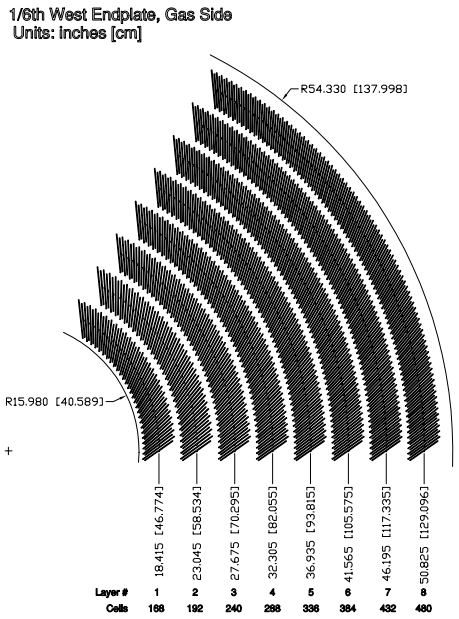


Figure 3.7: COT superlayers.

Central Outer Tracker

The COT [41] is a multi-wire open-cell drift chamber for charged particle reconstruction, occupying the radial region from 44 to 132 cm and $|z|$ <155 cm. COT uses small drift cells, ∼2 cm wide, and a fast gas to limit drift times to less than 130 ns. Each cell consists of 12 sense wires oriented in a plane, tilted with respect to the radial. A group of such cells at a given radius is called a superlayer. There are eight alternating superlayers (Figure 3.7) of stereo (nominal angle of 2°, used for r-z measurement) and axial (used for r-ϕ measurement) wire planes. The main parameters of the COT are summarized in Table 3.2.

The COT is filled with a mixture of Argon:Ethane = 50:50 which determines the drift velocity. A charged particle travels through the gas mixture and produces ionization electrons. The electrons drift toward the sense wires in the electric field created by cathode field panels and potential wires of the cell. In the crossed magnetic and electric fields electrons originally at rest move in the plane perpendicular to the magnetic field at an angle α, Lorentz angle, with respect to the electric field lines. The value of α depends on the magnitude of both magnetic and electric fields and the properties of the gas mixture. The optimal situation for the resolution is when the drift

Table 3.2: Design parameters of the CDF tracking systems

Layer 00	
Radial coverage	1.35 to 1.65 cm
Resolution per measurement	6 μm (axial)
Number of channels	13,824
SVX II	
Radial coverage	2.4 to 10.7 cm, staggered quadrants
Number of layers	5
Resolution per measurement	12 μm (axial)
Total length	96.0 cm
Rapidity coverage	$\|\eta\| \leq 2.0$
Number of channels	423,900
ISL	
Radial coverage	20 to 28 cm
Number of layers	one for $\|\eta\| < 1$; two for $1 < \|\eta\| < 2$
Resolution per measurement	16 μm (axial)
Total length	174 cm
Rapidity coverage	$\|\eta\| \leq 1.9$
Number of channels	268,800
COT	
Radial coverage	44 to 132 cm
Number of superlayers	8
Measurements per superlayer	12
Maximum drift distance	0.88 cm
Resolution per measurement	180 μm
Rapidity coverage	$\|\eta\| \leq 1.0$
Number of channels	30,240

direction is perpendicular to that of the track. To make the ionization electrons drift in the ϕ direction, all COT cells are tilted by $\alpha = 35^\circ$ with respect to the radial. The COT track resolution could be estimated from equation:

$$\frac{\sigma(p_T)}{p_T} = \frac{\sigma_{r\phi} P_T}{0.3BL^2}\sqrt{\frac{720}{N+4}} \tag{3.3}$$

where lever arm L=1 m, the magnetic flux density B=1.4 T, the $r-\phi$ hits resolution $\sigma_{r\phi} = 140\mu m$. In the case of the number of hits N=96, it's very close to the measured resolution performance $\delta p_T = 0.15\%/p_T$.

3.2.3 The Calorimeter

Energy measurement is done by sampling calorimeters, which are absorber and sampling scintillator sandwiches with phototube readout. Outside the solenoid, Pb-

scintillator electromagnetic (EM) calorimeters are designed to fully contain showers from electrons and photons. The Fe-scintillator hadronic (HAD) calorimeters surround EM calorimeters could measure energy loss of hadrons (π,k,n,p) by nuclear cascades. The central ($|\eta| < 1.1$) calorimeter systems have been retained from Run I, but the plug ($1.1 < |\eta| < 3.6$) calorimeters are new for Run II.

Both the central and plug electromagnetic calorimeters have fine grained shower profile detectors at electron shower maximum ($\sim 6X_0$), and preshower pulse height detectors at approximately one radiation depth ($1X_o$) . The region, chimney, $0.77 < \eta < 1.0$, $75^\circ < \phi < 90^\circ$ is un-instrumented to allow for cryogenic utilities servicing the solenoid.

Central Calorimeters

The central calorimeters consist of the central electromagnetic calorimeter (CEM) [42], the central hadronic calorimeter (CHA) [43], and the end wall hadronic calorimeter (WHA). To provide more accurate information on the position of the electromagnetic shower inside the calorimeter, the Central Electromagnetic Shower (CES) [42] detector is embedded inside the CEM at the shower maximum, at a depth of approximately 6 radiation lengths.

The CEM and CHA are constructed in wedges which span 15° in azimuth and extend about 250 cm in the positive and negative z direction, shown in Figure 3.8. There are thus 24 wedges on both the $+z$ and $-z$ sides of the detector, for a total of 48. A wedge contains ten towers, each of which covers a range 0.11 in pseudo-rapidity. Thus each tower subtends $0.11 \times 15^\circ$ in $\eta \times \phi$. The CEM covers $0 < |\eta| < 1.1$, the CHA covers $0 < |\eta| < 0.9$, and the WHA covers $0.7 < |\eta| < 1.3$. Table 3.3 shows important parameters for Calorimetry.

The CEM uses lead sheets interspersed with polystyrene scintillator as the active medium and employs phototube readout, approximately $19X_0$ in depth, and has an energy resolution of $13.5\%/\sqrt{E_T} \oplus 2\%$[1]. The CHA and WHA use steel absorber interspersed with acrylic scintillator as the active medium. They are approximately 4.5λ in depth, and have an energy resolution of $75\%/\sqrt{E_T} \oplus 3\%$, as measured on the test beam for single pions [43].

The CES detector is a proportional strip and wire chamber situated at a radius

[1] $\oplus$ denotes addition in quadrature

Table 3.3: Central and Plug Upgraded Calorimeter Comparison.

	Central		Plug
EM:	CEM		PEM
Coverage	$\|\eta\| < 1.1$		$1.1 < \|\eta\| < 3.6$
Thickness	$19X_0, 1\lambda$		$21X_0, 1\lambda$
Sample (Pb)	$0.6X_0$		$0.8X_0$
Sample(scint.)	5 mm		4.5 mm
Wedges	24		12
Towers/Wedge	10		12
Resolution	$13.5\%/\sqrt{E_T} + 1.7\%$		$16\%/\sqrt{E} + 1\%$
Shower Max. seg. (cm)	$1.4\phi\times$(1.6-2.0) Z		0.5×0.5 UV
Hadron:	CHA	WHA	PHA
Coverage	$\|\eta\| < 0.9$	$0.7 < \|\eta\| < 1.3$	$1.1 < \|\eta\| < 3.6$
Thickness	4.5λ	4.5λ	7λ
Sample (Fe)	1 to 2 in.	1 to 2in.	2 in.
Sample (scint.)	10 mm	10mm	6 mm
Resolution	$75\%/\sqrt{E_T} \oplus 3$	$75\%/\sqrt{E_T} \oplus 3$	$80\%/\sqrt{E} + 5\%$

of 184 cm from the beamline. In the azimuthal direction, cathode strips are used to provide the z position and in the ϕ direction, anode wires are used. These wires can effectively measure the transverse shower profile to distinguish between a single shower from a prompt photon and two showers from a decay of a neutral meson to two photons, e.g. $\pi^0 \to \gamma\gamma$, with a position resolution of 2 mm at 50 GeV. The minimum resolution of these two photons at the strip chamber is 50 cm GeV/c/p_T [44] to be compared with the channel spacing in this chamber of 1.45cm to 2.0cm. The Moliere radius of a lead plus scintillator calorimeter is 3.5 cm. The shower size of CES cluster is usually 11 strips $\times$ 11 wires.

The fractional energy observed for incident pions in CEM and HCAL is 40% and 20% at 2 GeV/c. It becomes plateau of 30% and 50% at 8 GeV/c [45]. The modeling of energy response for low momentum hadrons affect more on the $\not\!\!E_T$ resolution. As for high momentum hadrons, the energy response compared to electron is ~ 1.3. This affects the calorimeter energy resolution and makes CDF calorimeter non-compensating.

Plug Calorimeters

The plug calorimeters consist of the plug electromagnetic calorimeter (PEM) [46], and the plug hadronic calorimeter (PHA). At approximately $6X_0$ in depth in PEM is the plug shower maximum detector (PES). Figure 3.10 shows the layout of the detector and

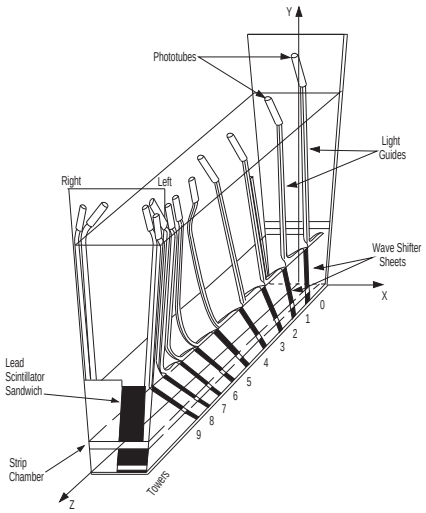


Figure 3.8: CEM/CES/CHA wedge.

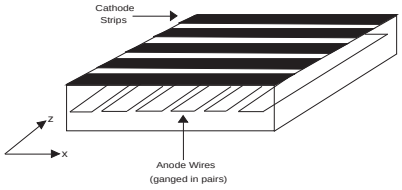


Figure 3.9: CES strip and wire.

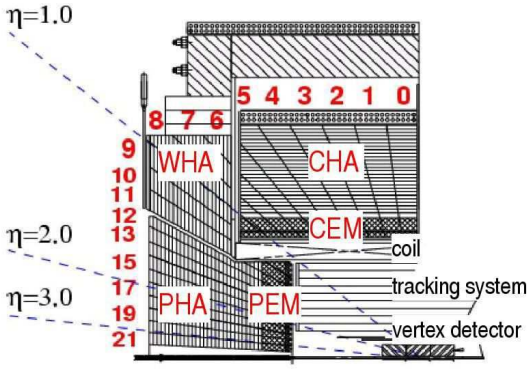


Figure 3.10: PEM/PES/PHA layout.

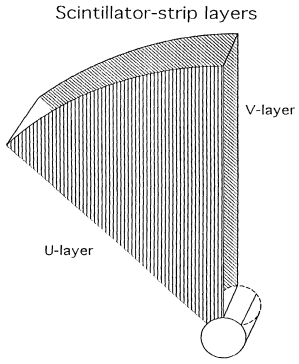


Figure 3.11: PES U and V layers.

coverage in polar angle $36.8° > \theta > 3°$ ($1.1 < |\eta| < 3.64$). Each plug wedge spans 15° in azimuth, however in the range $36.8° > \theta > 13.8°$ ($1.1 < |\eta| < 2.11$) the segmentation in azimuth is doubled and each tower spans only 7.5°.

The PEM is a lead-scintillator sampling calorimeter. It is approximately $21X_0$ in depth, and has an energy resolution of $16\%/\sqrt{E} \oplus 1\%$. The PHA is a steel-scintillator sampling calorimeter. It is approximately 7λ in depth, and has an energy resolution of $74\%/\sqrt{E} \oplus 4\%$, as measured on the test beam for single pions [43]. The PES consists of two layers of scintillating strips: U and V layers offset from the radial direction by +22.5° and −22.5° respectively, as shown in Figure 3.11. The position resolution of the

PES is about 1 mm.

3.2.4 Muon Systems

The muon with $P_T < 300$ GeV/c is a minimum ionizing particle which loses very little energy in detector materials. The muon's lifetime, 2.2 μs, is long enough to pass through all the detector components, reach the muon chambers and decay outside.

The major background of muon detection is hadrons produced in $p\bar{p}$ collisions. Hadrons are largely shielded by calorimeter and return yoke (Fe) through hadronic showering. The material budget for hadron calorimeter and muon system are 10λ in the barrel region. Figure 3.12 shows the absorption length of muon system as a function of η. It is equivalent to 1.7m thickness of iron and the punch-through probability for 50 GeV/c pion is 7.2% [47]. The punch-through has two components: a hadronic part which is a hadronic cascade shower particle leaking from the shield and penetrating muons from pion or kaon decay.

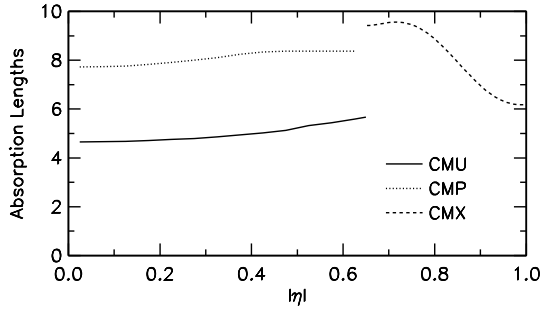


Figure 3.12: The interaction length of material before hitting muon systems

There are four independent muon systems [48]: the central muon detector (CMU), the central muon upgrade (CMP), the central muon extension (CMX), and the intermediate muon detector (IMU). The calorimeter steel serves as a filter for muon detection in the CMU and CMX, over the range $|\eta| < 1$, $p_T > 1.4$ GeV. Additional iron shielding, including the magnet yoke, provides a muon filter for the CMP in the range $|\eta| < 0.6$, $p_T > 2.2$ GeV. The (non-energized) forward toroids from Run I provide muon filters for IMU in the range $1.0 < |\eta| < 1.5$ for $p_T > 2$ GeV. Only CMU are kept from Run I and the other muon chambers are added along with scintillation counters near each detectors to identify the beam crossing that produced the muons: upgraded scintillator (CSP), extension scintillator (CSX) and barrel scintillator (BSU).

Table 3.4: Design parameters of the CDF II Muon Detectors.

	CMU	CMP	CMX	IMU
Pseudo-rapidity coverage	$\|\eta\| \leq 0.6$	$\|\eta\| \leq 0.6$	$0.6 \leq \|\eta\| \leq 1.0$	$1.0 \leq \|\eta\| \leq 1.5$
Drift tubes				
Cross-section, cm	2.68 x 6.35	2.5 x 15	2.5 x 15	2.5 x 8.4
Length, cm	226	640	180	363
Max drift time, μs	0.8	1.4	1.4	0.8
Scintillation counters				
Thickness, cm		2.5	1.5	2.5
Width , cm		30	30-40	17
Length , cm		320	180	180
Minimum muon p_T, GeV	1.4	2.2	1.4	1.4-2.0
multiple scatt. res., cm/GeV	$12/p$	$15/p$	$13/p$	$13 - 25/p$

The intermedium muon system is designed to trigger on muons with $|\eta| < 1.5$ and to identify off-line muons with $|\eta| < 2.0$. The heart of the detector is a barrel of CMP-like chambers and CSP-like scintillation counters mounted on the outer radius of the forward toroids. Unfortunately, the calibration and Monte Carlo simulation of this forward muon system is not ready for this thesis. However, the off-line forward muon could still be recovered by using forward tracks with minimum ionizing criteria.

A muon chamber contains a stacked array of drift tubes and operates with a gas mixture of Argon:Ethane = 50:50. The basic drift principle is the same as that of the COT, but the COT is a multi-wire chamber, whereas at the center of a muon drift tube there is only a single sense wire. The sense wire is connected to a positive high voltage (HV) while the wall of the tube is connected to a negative HV to produce a roughly uniform time-to-distance relationship throughout the tube. The drift time of a single hit gives the distance to the sense wire, and the charge division at each end of a sense wire can in principle be used to measure the longitudinal coordinate along the sense wire. The hits in the muon chamber are linked together to form a short track segment called a muon stub. If a muon stub is matched to an extrapolated track in the tracking system, a muon is reconstructed. The coverage for the muon systems in $\eta - \phi$ space is shown in Figure 3.13. CMU, CMP and CMX muon systems are also shown in Figure 3.14. The parameters for the muon systems are summarized in Table 3.4.

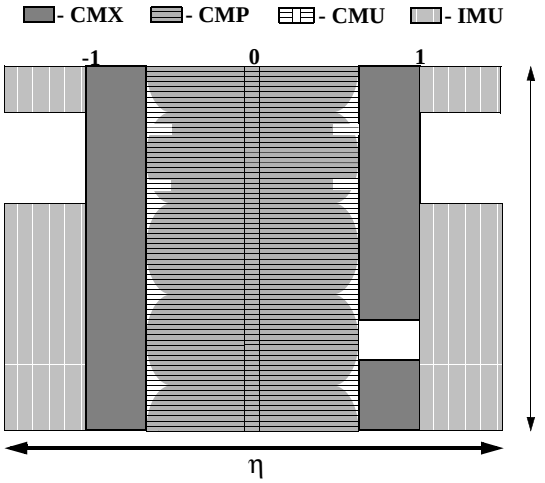


Figure 3.13: Location of the muon detectors in ϕ and η. On the east side, there is a gap in coverage in the CMX of 30° in azimuth, due to the location of the cryogenic utilities servicing the solenoid. The Miniskirt (labeld as M, 15° in azimuth) and Keystone (labeled as K, 30° in azimuth) of CMX along with IMU are not used in this analysis.

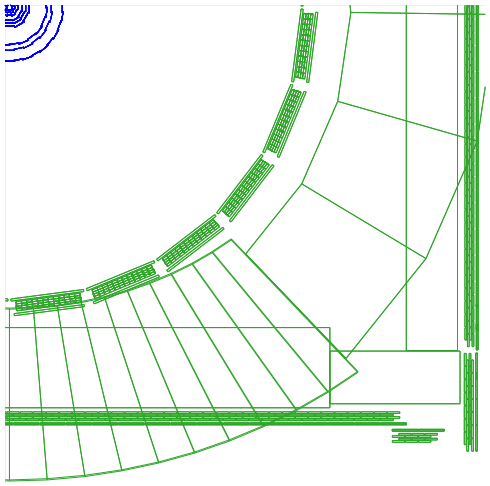


Figure 3.14: Central muon systems: CMU, CMP, CMX.

CMU and CMP

The CMU [49] is unchanged from Run I. It is located behind the towers of the CHA and divided into wedges covering 12.6° in azimuth for $\eta < 0.6$. Only muons with a $P_T > 1.4$ GeV reach the CMU. Each wedge has three towers, each comprised of four layers of four drift tubes. The second and fourth layers are offset by 2 mm in ϕ direction from the first and third.

A 50 μm diameter stainless steel resistive sense wire is located in the center of each cell. The wires in the cells in the first and third (second and fourth) layers are connected in the readout. Each wire pair is instrumented with a time-to-digital converter (TDC) to measure the ϕ-position of the muon and an analogue-to-digital converter (ADC) on each end to measure z position via charge division. The position resolution of the detector is 250 μm in the drift direction (r-ϕ) and 1.2 mm in the sense wire direction z.

Approximately 0.5% of high energy hadrons produced will pass through the

CMU creating an irreducible fake muon background. In order to reduce this effect, an additional muon chamber, the central muon upgrade (CMP), is installed behind 60 cm of steel. The CMP consists of a four-sided box placed on the outside of the CDF detector. It has single wire through aluminum extrusions, with cathode strips for field shaping. Muons with $P_T > 2.2$ GeV can reach the CMP. The rectangular form of the CMP detector means that its η varies in azimuth (Figure 3.13). The CMP covers $|\eta| < 0.6$.

The maximum drift time of the CMU is longer than the $p\overline{p}$ bunch crossing separation, which can cause an ambiguity in the Level 1 trigger (Section 3.2.6). To resolve the ambiguity scintillation counters are used. The scintillation counters (CSP) are installed on the outer surface of the CMP.

Central Muon Extension (CMX)

The CMX has eight layers and extends the η coverage to $0.6 < |\eta| < 1.0$. It consists of two 120° arches located at each end of the central detector, as shown in Figure 3.13. The un-instrumented regions have been filled by the insertion of a 30° keystone at the top, and two 90° miniskirts for the lower gaps. There is a gap in the coverage on the east side due to cryogenic utilities servicing the solenoid as shown in Figure 3.13, known as the "chimney".

A layer of scintillation counters (the CSX) is installed on both the inside and the outside surfaces of the CMX. No additional steel was added for this detector because the large angle through the hadron calorimeter, magnet yoke, and steel of the detector end support structure provides more absorber material than in the central muon detectors.

3.2.5 Luminosity Measurement

At the extreme forward region of the CDF-II detector, $3.75 < |\eta| < 4.75$, two modules of Cerenkov Luminosity Counters (CLC) [50, 51] are placed pointing to the center of the interaction region to record the number of $p\overline{p}$ interactions. The detector consists of low-mass gaseous Cerenkov counters with high light yield ($\sim$ 100 photoelectrons). The number of particles recorded in the CLC modules is combined with an acceptance of the CLC ($A \approx 0.6$) and the inelastic $p\overline{p}$ cross-section ($\sigma_{in} \approx 61\ mb$) to determine the instantaneous luminosity using the following equation:

$$L = \frac{\mu \cdot f_{BC}}{\sigma_{\mathrm{in}} \cdot \mathrm{A}} \tag{3.4}$$

where L is the instantaneous luminosity, f_{BC} is the rate of bunch crossings in the Tevatron and μ is the average recorded number of $p\overline{p}$ interactions per bunch crossing. The inelastic cross-section $\sigma_{\text{in}} = 60$ mb is measured in [52]. The rate of bunch crossing is the number of bunches (N_B) multiplied by revolution time 47.7 kHz. It allows precise luminosity measurements at peak instantaneous luminosities of $2 \times 10^{32} cm^{-2} s^{-}1$, corresponding to an average of ~ 7 interactions per bunch crossing.

3.2.6 Trigger and Data Acquisition

The Tevatron collides 36 proton bunches with 36 anti-proton bunches with a bunch spacing of 12 bunches each. The bunch structure of the Tevatron has 3 bunch trains of 12 bunches each. Between the bunch trains are "abort gaps" of 3.5μs each, and the actual bunch crossing rate is 1.7 MHz. The CDF-II data acquisition system can store data at a maximum rate of 18 Mb/s. With an average event size of 170 kb, this translates into an average event rate of 100 Hz. The total hadronic cross-section (including the elastic, inelastic hard scattering, and inelastic diffractive processes) is about 80 mb [52]. However, interesting physics process, e.g. $b\overline{b}$ cross-section, is about 1000 times smaller, 0.1 mb. Extracting the most interesting physics events from the large number of events reduces the cost and time to reconstruct data.

This section is based on the most up-to-date description of CDF-II trigger in [53]. The CDF trigger system has a three-level architecture: Level 1 (L1), Level 2 (L2), and Level 3 (L3). The data volume is reduced at each level, which allows more refined filtering at subsequent levels with minimal deadtime. The decision time for L1, L2 and L3 is about 5.5 μs, 20 μs and 1 s, respectively. The event accept rate for L1, L2 and L3 is 50 kHz, 300 Hz and 30 Hz. The delay necessary to make a trigger decision is achieved by storing detector readout information in a storage pipeline, as shown in Figure 3.15.

Figure 3.16 shows the three level pipelined and buffered trigger systems. The original specification for the CDF-II detector allowed for operation with a Tevatron bunch spacing of 132 ns. The base CDF-II clock period utilized by the entire CDF trigger system (CDF_CLOCK) is 132 ns. To allow time for transmission and processing of the trigger signals, there is a 5.5 μs Level-1 latency from $p\bar{p}$ collision to Level-1 trigger decision. This requires each detector element to have local data buffering for 42 colock cycles.

At L1 axial layers of the COT are used by eXtreme Fast Tracker (XFT) [54]

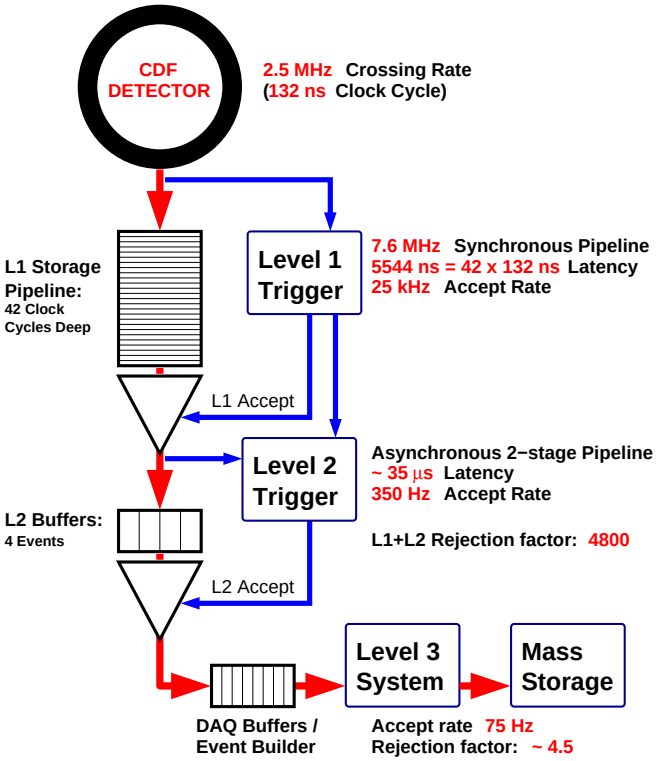


Figure 3.15: The CDF-II trigger and data acquisition system. Data is acquired at the beam crossing period of 396ns into a synchronous pipeline that is clocked at 132ns. The Level-1 decision is produced after 42 clock cycles, at which point event processing becomes asynchronous. Typical trigger rates and rejection factors for the three-level system are shown in the figure.

to reconstruct ϕ and P_T for the tracks. Based on the XFT tracks and a ratio of the hadronic energy to the electromagnetic energy of a calorimeter tower (HAD/EM ratio) electrons and photons are then reconstructed. Muons are reconstructed by matching XFT and muon hits. Jets are reconstructed based on a sum of the electromagnetic and hadronic energies for a tower. $\not\!\!E_T$ and $\sum E_T$ (a scalar sum of the energies of all of the calorimeter towers) are also reconstructed at L1.

The Track Extrapolation System (XTRP) is a fully digital system that is utilized in the track-based selection of high momentum lepton and heavy flavor signatures.

At the first level (Level 1), a trigger decision is made based only on a subset of the detector and quick pattern recognition or simple counting algorithms. The second level of the trigger (Level 2) does a limited event reconstruction. The third level of the trigger (Level 3) uses the full detector information to fully reconstruct events in a

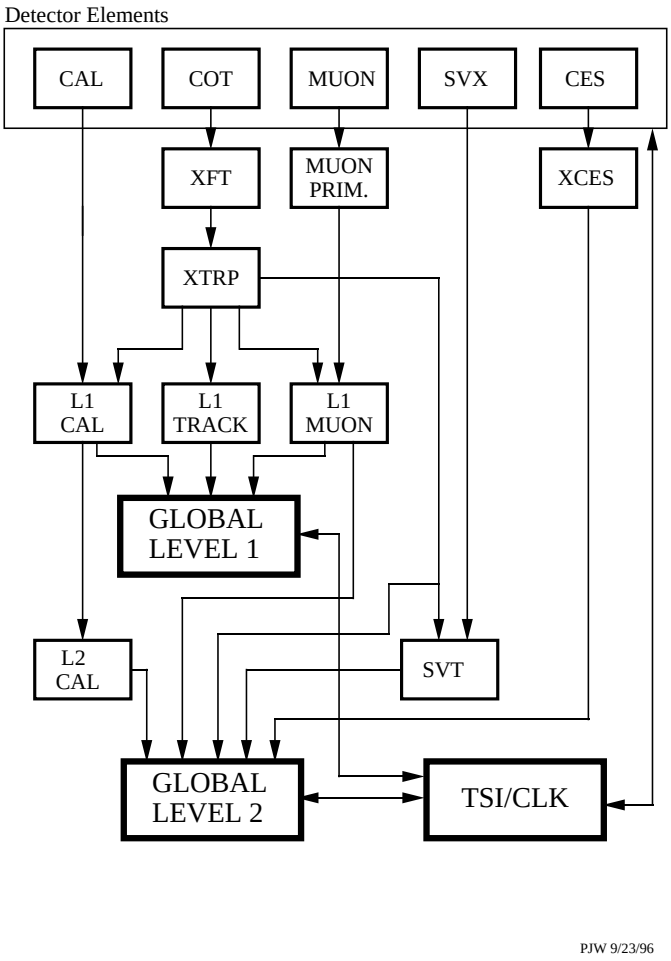


Figure 3.16: The CDF-II trigger system. Trigger primitives are acquired from the detector elements and lead to a Level-1 decision. In the case of the track-based triggers, the XFT finds tracks in the COT, which are passed to the XTRP system. In the XTRP, the tracks are extrapolated to the muon and calorimeter systems for muon and electron identification. Tracks are also passed onto the Track Trigger (L1 Track) by way of the XTRP. The XTRP additionally provides tracking information for the Silicon Vertex Trigger (SVT) and Level-2 trigger processor.

processor farm. Each sub-detector generates primitives which can be used in the trigger system to select events. The L1 trigger reduces the event rate from 1.9 MHz to 50 KHz, which is limited by the L2 processing time. The L2 decision time is about 20 μs. L2 is a combination of hardware and software triggers and is asynchronous. The L2 accept rate is about 300 Hz which is limited by the speed of the event builder in L3. L3 is a

purely software trigger consisting of the event builder running on a large PC farm. The L3 accept rate of about 75 Hz is limited by the speed of writing data to tape for the permanent storage.

3.3 CDF Computing

Computing plays an important role in high energy physics because of the increasing need for parallel offline data processing and large data transfer across the internet. UCSD group has designed, commissioned and operated a powerful system to make distributed computing resources accessible from desktop and laptops anywhere. We call it the CDF Analysis Facility, or CAF.

Since the beginning of Run II in 2001 the Fermilab Tevatron has been delivering proton-antiproton collisions to CDF with a steadily increasing performance. By December 2007, Run II gathered more than $2.5 fb^{-1}$ of data, equivalent to 3.6×10^9 events or 1070 TBytes of raw data, and this is expected to double during the next year. By the end of Run II (2010) up to 8 fb^{-1} of integrated luminosity are expected to be available on tape.

The management of computing access at CDF is based on the concept of CDF Analysis Farms (CAFs) [55]. It is one of the first realization of large computing before the GRID technique is matured. In this model, users can submit jobs to specified computing farms, assign job output locations and monitor on-going jobs interactively. The data handling of storage element at CDF is managed by the SAM [56] system. In this model, a central database system manages all the experiment files and delivers them to the execute nodes. The use of caching and optimization make the file delivery very efficient.

3.3.1 Central Analysis Farm (CAF)

The CDF Analysis Farm(CAF) was designed and assembled in 2002, originally as a PC cluster localized at FNAL. Successively, the CAF model was exported offsite and several decentralized CDF Analysis Farms (dCAF) arose in many sites worldwide. In 2006, $\sim$50% of CDF computing power was outside Fermilab as shown in table 3.5. Since data are stored at Fermilab, the FNAL CAF is mostly used for data analysis jobs, whereas the dCAFs are mainly devoted to MC production. There are some remote sites (CNAF) that allow users to run analysis jobs on dataset replicas.

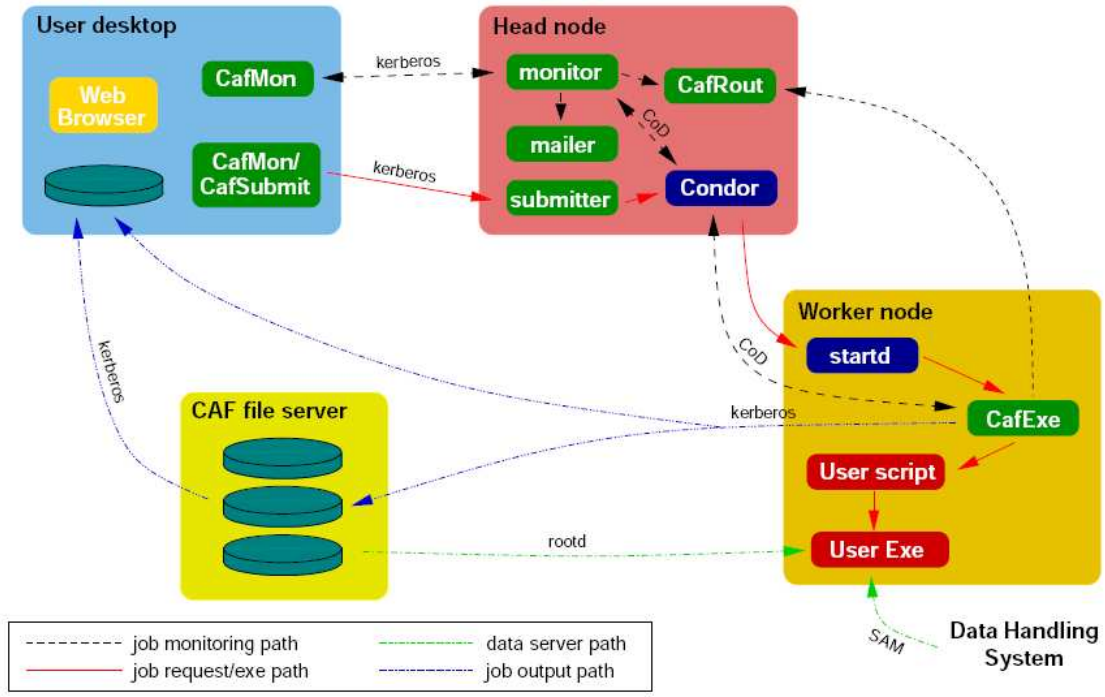


Figure 3.17: Schematic block of CAF. The data flow, command lines and authentication paths are indicated. The central batch system is Condor [57].

The CAF realization was a large batch-based cluster of commodity PC hardware: dual Intel/AMD processors and IDE disk, configured in RAID 50 arrays. CAF nodes are functionally divided in one head node, where the control and management daemons run, and many worker nodes, on which user jobs are executed. All nodes run Linux as an operating system and have access to the CDF software.

The farm batch manager is Condor [57], configured with just a submission mode, which plays also the role of management node and many worker nodes. Six batch slots, so called virtual machines (VMs), are allocated on every worker node, for each slot a generic user account is created. More details on Condor configuration can be found elsewhere. Figure 3.17 schematized the elements which make up the CAF: In User desktop, one user runs the user front-end applications for jobs submission (CafGui and CafSubmit), which sends the user submission parameters to the head node, and the user interactive tool to manage and monitor the running jobs (CafMon). CafGui also creates the executable tarball. Both communicate with the headnode via a kerberos authenticated connection. A web browser is used to access the web monitoring pages. The headnode hosts al the CAF and Condor management processes. The submission

Table 3.5: CAF's resources around the world (2006)

Cluster Name	Location	CPU [GHz]	Disk [TB]
CAF	FNAL	3200	370.0
TORCAF	Toronto (Canada)	576	10
CNAFCAF	CNAF (Italy)	480	32
MITCAF	MIT (USA)	322	3.2
SDSCCAF	San Diego (USA)	300	4.0
KORCAF	KNU (South Korea)	178	5.1
JPCAF	Tsukuba (Japan)	152	10.0
ASCAF	Academic Sinica (Taiwan)	134	3.0
HEXCAF	Rutgers (USA)	100	4.0
CANCAF	Cantabria (Spain)	50	1.5
Total		5492	442.8

request manager (submitter) grants kerberos authentication to users, transfers the CAF tarball from the user desktop, created the Condor submission files and submits the CAF job wrapper (CafExe) to Condor.

The monitoring daemon (monitor) provides the user with monitoring information and manages the user interactive requests. The CAF router (CafRout) allows communication between the monitor daemon and CafExe. The notification daemon (mailer) sends a summary mail upon job completion and cleans up the Condor files. Condor starts CafExe, which gets the user kerbeors credentials, untars the user tarball, and runs the user initial command. At the end of the job execution it archives the working directory and transfers the tarball to the output location. Moreover, CafExe performs some monitoring tasks: creates job summary file and serves as interactive CafMon callback. The user executable can either get data from the data handling system or read data from the CAF file server via rootd. All outward connections are kerberized.

3.3.2 Data Handling System - SAM

The CDF data is managed by the SAM [56] system, centered around a central database system that contains the metadata and the location of all the experiment files. The files at any location are managed by a set of services, called a SAM station. Each SAM station has a storage area associated with it that manages autonomously.

Files are grouped in datasets, that are generated both by the production tools and by final users themselves. A whole dataset can be used to create a SAM project, that will be used to access data via a specified SAM station. Once a project is created,

processes can ask for files, and the SAM station will serve files starting with the ones already in the local storage, and fetching the others from other SAM stations in parallel, thus minimizing the wait times.

Using this mechanism, CDF users are accessing over 1.5PBytes of data. Fermilab hosts both the tape library that contains all the data and a 370 TBytes disk cache. The second largest SAM station is hosted at the Italian CNAF, with a 32 TBytes disk cache. While CDF has done well with delivering files to the jobs, the handling of job output was left to the users. The CAF itself does send back user log files and smaller data files, but users are still required to manage the bulk of their data by themselves.

Chapter 4

Data Sample and Event Selection

In this chapter, the data sample and candidate selections are described. The trigger criteria for the datasets are based on a combination of the signatures of a single, high P_T lepton and large missing transverse energy. A brief introduction of corrections of low-level objects, i.e. tracks, electrons, muons, jets and $\not\!\!E_T$ will be described systematically. The high-level cuts to select good quality candidates with as large an acceptance as possible are presented. The general strategy for the candidate selection is to apply loose cuts for the suppression of dominant backgrounds and to rely on the multivariate analysis to discriminate the remaining backgrounds from the signal.

4.1 Trigger Paths and Datasets

The signature of a dilepton+$\not\!\!E_T$ event contains special characteristics that can be used to separate them from dominant background events. High-P_T (> 10 GeV/c) lepton requirement reduces soft leptons originating from heavy flavor, e.g. lepton from semileptonic deacy of $b(c)$ quark or from charged hadron (π/K) which mimic the lepton. Large missing transverse energy ($\not\!\!E_T > 15$ GeV/c) rejects backgrounds have no neutrino in the final states, e.g. Drell-Yan process (Z/γ). At trigger level, we attempt to select signal events based on these two signatures with high efficiency while keeping the trigger accept rate low.

To meet these criteria, four trigger paths are used in this analysis to filter collision data based on lepton types and detector regions. We have two electron trigger paths, ELECTRON_CENTRAL_18 and MET_PEM. ELECTRON_CENTRAL_18 is a central electron trigger which requires at least one high P_T (> 18 GeV/c) track pointing to a

central calorimeter shower maxima. MET_PEM uses the high $\not\!\!E_T$ ($\not\!\!E_T > 15$ GeV) signature and requires at least one high E_T (> 20 GeV) EM cluster in the plug detector. There are no track requirements due to the poor tracking coverage in the forward region. We have two central muon triggers, CMUP_CENTRAL_18 and CMX_CENTRAL_18. Both of them require at least one high P_T (> 18 GeV/c) track pointing toward a reconstructed muon stub.

The combined averaged trigger accept rate is 1.25 Hz which occupies 1.7% of the total 75 Hz bandwidth at Level 3. The trigger efficiency ranges from 90% to 97% for different trigger path and is measured from W-rich or Z-rich data described in detail in Appendix A.

A trigger path is a series of trigger criteria for three selection stages, L1, L2, and L3. The sequential criteria for passing each of the three stages of each trigger path are described below.

- ELECTRON_CENTRAL_18

  The central electron trigger uses tracker and calorimeter information. The average trigger accept rates during the data collection period were 64 Hz, 2 Hz and 1 Hz at Level 1, 2 and 3, respectively. The 1 Hz event rate gives a cross-section of 50nb and the trigger efficiency is measured to be $96.2 \pm 0.6\%$.

  - Level 1: A central EM trigger tower ($|\eta| < 1$) with $E_{TEM} \geq 8$GeV. For clusters with energy less than 14 GeV, E_{Had}/E_{EM} must be less than 0.125 to reject hadronic particles. The trigger tower containing the cluster must have a matched XFT track with $P_T \geq 8$GeV/c.
  - Level 2: A central EM cluster with $E_{TEM} \geq 16$GeV and $E_{Had}/E_{EM} \leq 0.125$ while a cluster is formed by adding adjacent towers with $E_T > 7.5$GeV. The trigger cluster must have a matched XFT track with $P_T \geq 8$GeV/c
  - Level 3: A central EM cluster with $E_{TEM} \geq 18$GeV and $E_{Had}/E_{EM} \leq 0.125$. The seed tower of the EM cluster must be matched to a fully reconstructed 3D track with $P_T \geq 9$GeV/c

- MET_PEM

  The COT coverage only extends to $|\eta| < 1.5$ and the PEM detector coverage is $1.2 < |\eta| < 2.5$. Therefore, we only require high E_T electromagnetic shower and

large $\not\!\!E_T$ criteria to trigger the signal events. The averaged trigger acceptance rate for Level 1, 2, 3 were 49 Hz, 1 Hz and $\sim$ 0.9 Hz, respectively. The averaged efficiency of this trigger is $91.0 \pm 0.1\%$ which actually has a slow turn-on curve starting from $\not\!\!E_T = 16$GeV and reaching plateau at 99% after $\not\!\!E_T > 30$GeV.

- Level 1: Forward EM cluster with $E_{TEM} \geq 8$GeV and $E_{Had}/E_{EM} \leq 0.125$ for clusters with energy less than 14 GeV. The $\not\!\!E_T$ must be more than 15 GeV (z=0 assumed).
- Level 2: Forward EM cluster with $E_{TEM} \geq 20$GeV and $E_{Had}/E_{EM} \leq 0.125$ for all clusters.
- Level 3: Forward EM cluster with $E_{TEM} > 20$GeV and $E_{Had}/E_{EM} \leq 0.125$ for all clusters and $\not\!\!E_T > 15$GeV (z=0 assumed).

- MUON_CMUP18

  The central muon trigger utilizes information from both the CMU and CMP ($|\eta_{det} < 0.6|$). A high P_T track is required to match to the muon stubs. The trigger accept rate were 30 Hz, 4 Hz and 0.15 Hz at Level 1, 2 and 3, respectively The efficiency is measured to be $89.8 \pm 0.5\%$.

  - Level 1: A CMU stub with $P_T \geq 6 GeV/c$ matched to a XFT track with $P_T \geq 4.09 GeV/c$ and a stub in the CMP.
  - Level 2: No requirement.
  - Level 3: A fully reconstructed 3D COT track with $P_T \geq 18$GeV/c and matched to the CMU stub with $|\Delta X|$ of 10 cm and within $|\Delta X|$ of 20 cm of the CMP stub.

- MUON_CMX18

  The extended central muon trigger uses CMX which covers the pseudo-rapidity range $0.8 < |\eta_{det}| < 1.0$. The trigger accept rate were 2 Hz, 2Hz and 0.1 Hz at Level 1, 2 and 3, respectively. The efficiency is measured to be $96.7 \pm 0.4\%$.

  - level 1: a CMX stub with $P_T \geq 6$GeV/c. Some hits in the central muon wall scintillator extension (CSX) detector and a COT track with $P_T \geq 8.34$GeV/c passing through at least 4 super layers.

Table 4.1: Luminosity for each of the good run lists (version 17) used in this analysis. A factor of 1.019 has been applied to each number to present the absolute accumlated luminosity.

Good run list	$\mathcal{L}$ (pb^{-1})
EM_NOSI	1906.1
EM_CMUP_NOSI	1876.6
EM_MU_NOSI_CMXIGNORED	1805.5
EM_SI	1789.5
EM_CMUP_SI	1762.9
EM_MU_SI_CMXIGNORED	1695.4

- level 2: No further requirements.
- level 3: A fully reconstructed 3D COT track with $P_T \geq 18$GeV/c matched to the CMX stub to within 10 cm in r-ϕ plane.

Dataset

There are three data sets associated with the four trigger paths. Due to the fact that not all the detectors are operated normally, there are six good run lists defined to indicate the detector subsystems that are functional at that time. The six good run lists and their corresponding luminosity are shown in Table 4.1. The meaning of the labels are described below:

EM means the calorimeter was working. Every event is required to have good EM and Hadron calorimeter conditions because of the $\not\!\!E_T$ reconstruction and the minimum ionizing requirement for muon identification. **SI** and **NOSI** mean silicon detectors are operational or not, respectively. Good silicon is only required for the forward lepton types: PHX and CMIOPES. **CMUP** means both CMU and CMP are in working order. **CMXIGNORED** means early runs (about 71 nb^{-1}) with bad CMX were ignored. It is only used when one of the lepton pairs is a muon detected by CMX detector.

4.2 Low-level Objects

This section describes how the low level objects (tracks, leptons, jets, and $\not\!\!E_T$) are reconstructed. The identification cuts are defined to suppress other kinds of particles mis-identified as leptons (fakes). The lepton identification efficiencies are measured from

Z-pole data through tag-and-probe methods and described in Appendix B. The probability of a lepton to be mis-identified from a jet (fake) is measured from a subsample of jet-rich data contains lepton-like objects and is described in Appendix C.

There are three dominant sources of fake electrons: a) photons from π^0, η and other mesons that convert into asymmetric e^+e^- pairs in the material before the COT volume, b) charged hadrons in jets that either interact in the electromagnetic volume of the calorimeter through charge exchange $\pi^- p^+ \rightarrow \pi^0 n$ or overlap with a π^0 or secondary photon in the jet, and c) electrons from the decay of heavy flavor (b, c). The contribution from pion or kaon leptonic decay is small due to the helicity suppression. The photon conversions are suppressed by conversion algorithm which looks for two opposite sign tracks around candidate electrons with $|\Delta xy|$ <0.2 cm and $|\Delta cot(\theta)|$ <0.04. The hadronic fake electrons are suppressed by imposing isolation cut which will be defined later.

The fake muons could come from the decay of heavy flavor (b, c) in an energetic jet. It could also come from a pion or kaon decays before the COT volume results in a muon whose momentum is correctly measured. A pion or kaon which decays in flight inside the COT volume may cause a "kink" in the track which can be exploited to suppress this background by making a requirement on the track fit χ^2. A non-interacting punch-through hadron can be mis-identified as a muon. The suppression of muon fakes mainly rely on the isolation cut.

4.2.1 Lepton Identification and Categories

There are seven lepton types used in this analysis. They are categorized based on lepton flavor and the detector region. Almost all the fiduciality cuts are based on the tracks, thus allowing us to make non-overlapping categories. Among the lepton types there are two types of electrons, Tight Central Electron (TCE) and Phoenix Electron (PHX), and four types of muons, Central Muon (CMUP), eXtended Muon (CMX), Central Minimal Ionized Particle (CMIOCES) and Forward Minimal Ionized Particle (CMIOPES). The last type of lepton is crack track (CrkTrk) which is a lepton not fiducial to electron nor muon detector and considered as a muon or an electron candidate. The CMIOPES and CrkTrk are new lepton categories by comparing to previous CDF analysis [5] [4].

Track Reconstruction

The electron or muon coming from the leptonic decay of a W boson is expected to produce a track with high transverse momentum. Five of the lepton types- TCE, CMUP, CMX, CMIOCES, CrkTrk, require a well-reconstructed central track. The main detector used in the reconstruction of these tracks is the COT. The track reconstruction is performed as follows. Track segments are formed from hits recorded by the sense wires on each individual COT superlayer. Two separate algorithms are then used to form axial-only tracks: the segment finding algorithm links track segments between axial superlayers while the histogram linking algorithm performs a circle fit that adds axial hits starting from a seed track segment. Tracks found by both algorithms are merged together. Hits in stereo layers are then added using a full three-dimensional fit of the helical trajectory of the track. Finally a fit is performed where the tracks are forced to originate from the beamline in the x-y plane to improve the momentum resolution (beam constraint fit). The COT-only track reconstruction efficiency is 98% for $P_T > 5$ GeV/c and $|\eta| < 0.9$. Figure 4.1 shows the track reconstruction efficiency as a function of η. The COT-only track efficiency drops to zero rapidly at $|\eta| > 1.2$. In the forward region up to $|\eta| < 1.5$, we also used silicon seeded tracks extrapolated to COT (IO tracks). The silicon standard alone (SISA) is not used at all. The tracks used by CMIOPES are reconstructed forward tracks which require combined hits from the COT and SVX.

There is a separate type of track, PHX track, used in the selection of forward electron, PHX. It is a calorimeter seeded track and will be described later in the PHX category.

Tight Central Electron (TCE)

Constructing an electron object in the central region of the detector requires three objects: a cluster in the EM calorimeter, a track pointing to that cluster, and a shower-max cluster in the CES associated with the track. Table 4.2 lists the identification cuts of central electrons. All items are described below:

- conversion: The COT track is required to be inconsistent with a conversion $\gamma \rightarrow e^+e^-$. The conversion-finding algorithm looks for a conversion partner track that is opposite signed, originates from a common point (r-ϕ separation $\Delta XY < 0.2$ cm) and is parallel to the COT track ($\Delta cot\theta < 0.04$). The prompt electron often has a collinear bremsstrahlung photon, and this photon may convert in the detector

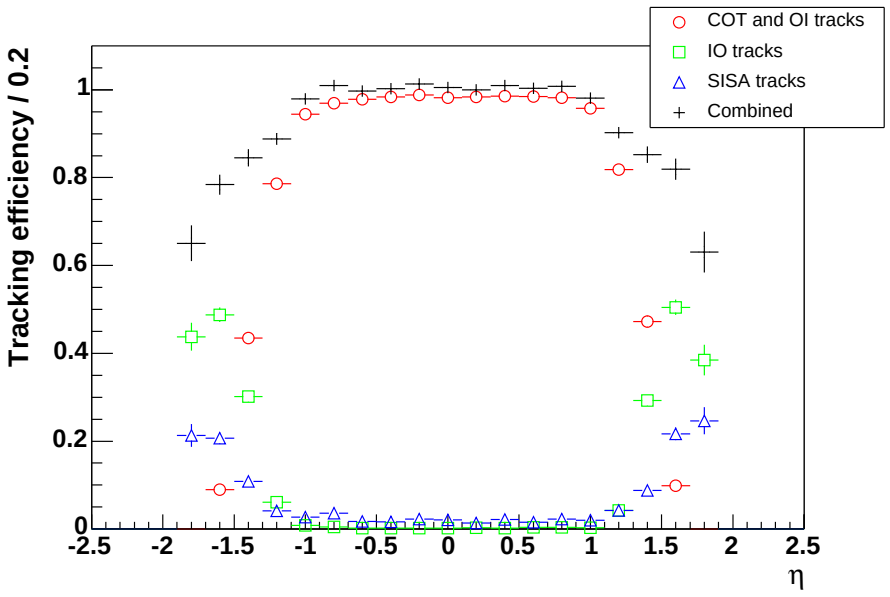


Figure 4.1: The track reconstruction efficiency as a of function of η. Outside-In (OI): COT tracks seeded tracks include SVX hits. Silicon Standard Alone (SISA): silicon hits only tracks.

material and produce one or two soft tracks. To avoid rejecting this type of electron ($e^{\pm} \rightarrow e^{\pm}\gamma$), the same algorithm is applied to the partner track to search for a third partner track. This type of lepton is called a "trident" and is accepted for analysis. Figure 4.2 shows the N-1 distribution for conversions.

- Had/Em: This is the ratio of total energy deposited in the EM calorimeter and Hadron calorimeter. It is used to reject candidates that arise from hadrons faking an electron. The energy dependence of this cut allows a high efficiency for real electrons over a large energy range.

- Iso/E_T: The isolation energy is the transverse energy inside a cone defined by $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \leq 0.4$ centered on the EM cluster minus the transverse energy of the cluster, $E_T^{cluster}$.

- TIso/P_T: The track isolation is defined in similar manner as the total transverse momentum of all tracks other than the electron track and within a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \leq 0.4$. The minimum P_T of track is 0.315 GeV/c. For tracks with $P_T > 10$ GeV/c, Δz_0 between tracks and electron tracks is required to be less than

Table 4.2: Lepton identification cuts for Tight Central Electron (TCE).

	TCE
Region	Central
Fiducial	Track Fiducial to CES
Track p_T	≥ 10(5 if $E_t < 20$)
Track $\|z_0\|$	≤ 60 cm
#Ax SL (5hits)	≥ 3
#St SL (5hits)	≥ 2
Conversion	$!= 1$
Had/Em	$\leq 0.055 + 0.00045E$
Iso/E_T	≤ 0.1
Lshr	≤ 0.2
E/P	$< 2.5 + 0.015 * E_T$
Signed CES ΔX	$-3 \leq q\Delta X \leq 1.5$ cm
$CES\|\Delta Z\|$	< 3 cm
Tracks	BcTrk (Larry's Correction if Data)

5 cm.

- Lshr: The lateral profile of the shower as measured in the CEM and CES is required to match profiles of electromagnetic showers measured in test beam data [42]. The equation to calculate Lshr is:

$$L_{shr} = 0.14 \sum_i \frac{E_j^{adj} - E_i^{expected}}{\sqrt{(0.14\sqrt{E})^2 + (\Delta E_I^{expected})^2}}, \tag{4.1}$$

  where E_j^{adj} is the measured energy in the CES cell and the $E_j^{expected}$ is the expected deposit energy while $\Delta E_I^{expected}$ is the expected energy uncertainties.

- E/P: The ratio of the cluster transverse energy to track P_T must be less than 2. This further suppresses hadronic particles faking the electron signature.

- Signed CES Δ x and CES Δ z: The difference between x (r-ϕ) as measured in the CES and x as measured by extrapolating the track out to the CES. This is applied to identify electrons that have emitted bremsstrahlung radiation and consequently have a CES cluster in consistent with the track trajectory, and also to reject jets faking electrons. The asymmetry of the $q\Delta x$ cut is designed to keep bremsstrahlung radiation but reject jet fake electrons.

The rejection efficiencies of the identification cut we concern after applying all the rest of cuts, a.k.a. N-1 rejection efficiency, is measured for each selection variable. They

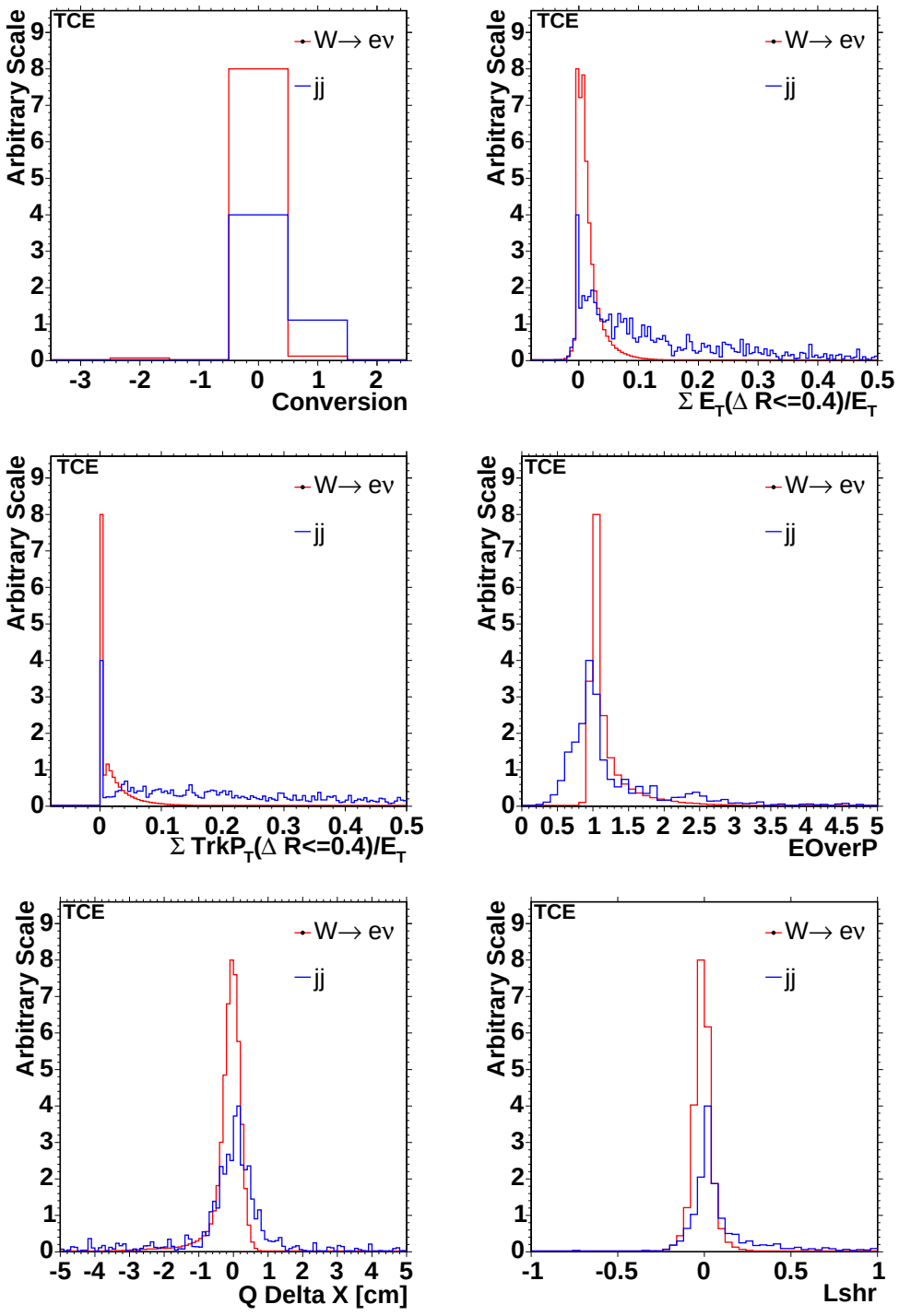
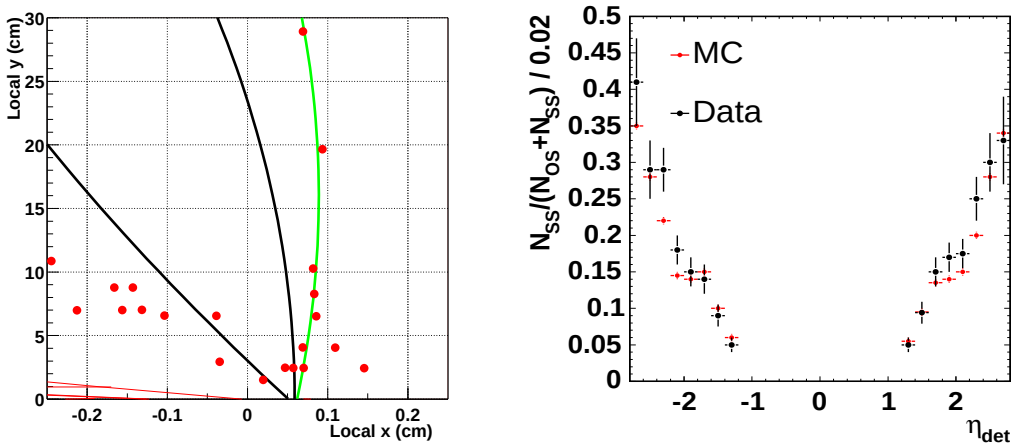


Figure 4.2: The N-1 distributions of identification variables for TCE for $W \rightarrow e\nu$ Monte Carlo and dijet Monte Carlo with $\hat{p}_T > 18$ GeV/c.

are 4%, 2%, 1% and 1% for conversion veto, isolation, E/P and $q\Delta x$, respectively. The TCE identification efficiency varies between $82.0 \pm 0.5\%$ and $76.3 \pm 0.7\%$ as luminosity increases.

Phoenix Tracking Forward Electron (PHX)



(a) (b)

Figure 4.3: (a) shows two calorimeter seeded tracks (black) and the PHX track (green) after attached silicon hits. (b) shows the charge mis-identification rate measured from TCE-PHX pair in the Z-pole.

Construction of a forward electron candidate object follows the same logic as in the central case, but with PEM and PES clusters rather than CEM and CES clusters. There is no track in the COT to use to match the PES cluster to the PEM cluster. Instead, the matching simply uses the highest energy PES cluster in the seed tower. For every Plug EM shower, we can calculate two hypothesis tracks with opposite sign charges and curvature $1/E_T$. These two hypothesis tracks are seeded from PES centroid and the event vertex as shown in the two black curves in figure 4.3. The silicon hits are associated along the PHX tracks. In this example, there is only one track associated with silicon hits (green curve). In the case of two PHX tracks are found, the track which has minimum χ^2/n.d.f. is selected [58].

The identification of plug electron starts from requiring a PEM 3x3 Fit tower in the fiducial region $1.2 < |\eta_{det}| < 2$. Table 4.3 shows the identification cuts for PHX. The explanations of each cut is listed below:

Table 4.3: Lepton identification cuts for phoenix electrons (PHX).

	PHX
Region	Plug
Pes2DEta	$1.2 < \|\eta\| < 2$
Had/Em	≤ 0.05
PEM3x3FitTower	true
$PEM3x3\chi^2$	≤ 10
Pes5x9U	≥ 0.65
Pes5x9V	≥ 0.65
Iso/Et	≤ 0.1
ΔR(Pes,PEM)	≤ 3.0
Track Match	True
NSiHits	≥ 3
Track $\|Z0\|$	$\leq 60cm$

- The Had/EM is the same definition as for the central electron and is required to be less than 0.05.
- PEM 3x3 Fit χ^2 compares the EM shower profile of the PEM cluster with test beam data.
- PES 5x9 U/V: The ratio of the energy in the central 5 strips of the PES cluster to the energy of the whole cluster of 9 strips. It is intrinsically an isolation ratio variable.
- PEM χ^2 rejects 8% of signal. The other cuts have 1.5% to 1% cut efficiency.

The PHX identification efficiency varies between $85.0 \pm 0.5\%$ and $79.4 \pm 0.6\%$ as function of time. The track finding efficiency is $86.7 \pm 0.5\%$. Figure 4.4 shows the N-1 efficiency for PHX identification variables.

Tight Central Muon (CMUP/CMX)

The muon candidate is formed by matching a COT track to the stubs in the muon chambers. The stubs are reconstructed as line segments formed by hits in the CMU, CMP, and CMX. The extrapolated track is required to match the position of the stub in the r-ϕ plane within 3, 5 and 6 cm in the CMU, CMP and CMX, respectively. The effect of multiple scattering is taken into account in the size of these cuts. The energy deposited in the calorimeters by the candidate track is required to be consistent with that of minimum ionizing particles to remove hadronic particles faking muons. A

Table 4.4: Lepton identification cuts for central muons: CMUP and CMX. The parts of CMX detector is defined in terms of track $\eta - \phi$.

	CMUP/CMX
CMU Fid	x-fid$<$ 0 cm z-fid$<$ 0cm
CMP Fid	x-fid$<$ 0 cm z-fid$<$-3cm
CMX Fid	x-fid$<$ 0 cm z-fid$<$-3cm
E_{em}	$\leq 2 + max(0, (p - 100) * 0.0115)$
E_{had}	$\leq 6 + max(0, (p - 100) * 0.028)$
Iso/Pt	≤ 0.1
NAxL(5 hits)	≥ 3
NStL(5 hits)	≥ 2
Track $\lvert Z_0 \rvert$	$\leq 60cm$
Track $\lvert D_0 \rvert$	$\leq 0.2cm$ (0.02cm if NSiHit$>$ 0)
χ^2/dof	≤ 4 (3 if run $\leq$ 186598)
$\lvert \Delta X_{CMU} \rvert$	$\leq 7cm$
$\lvert \Delta X_{CMP} \rvert$	$\leq 5cm$
$\lvert \Delta X_{CMX} \rvert$	$\leq 6cm$
ρ_{COT}	$>$ 140 cm if CMX
CMP veto	No Bluebeam in CMP for run $<$ 154449
CMX veto	No CMX for run $<$ 150144, No Miniskirt, No Keystone
Arches	Arches only for all run range
	Arches removing wedge 14 on West Side for run $>$ 190697
Tracks	BcTrk (Larry's Correction if Data)

CMP:	Bluebeam	$45^\circ < \phi < 52.3^\circ$
CMX:	Arches	$0^\circ \leq \phi \leq 75^\circ$ or $105^\circ \leq \phi \leq 225^\circ$ or $315^\circ \leq \phi \leq 360^\circ$
	Arches Wedge 14	$210^\circ < \phi \leq 225^\circ$ and $\eta < 0$
	Keystone	$75^\circ < \phi < 105^\circ$ and $\eta < 0$
	Miniskirt	$225^\circ < \phi < 315^\circ$

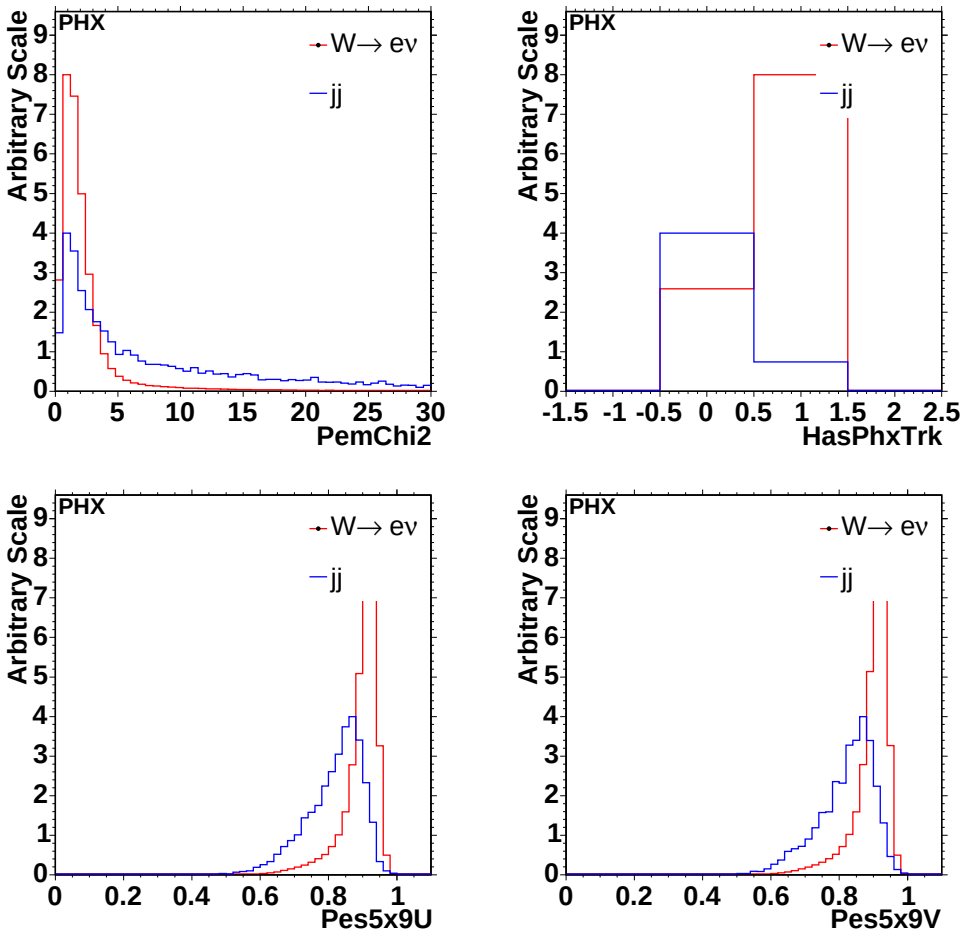


Figure 4.4: The N-1 distributions of identification variables for PHX.

minimum 2 GeV and 6 GeV energy is required at ECal and HCal. The energy isolation is also applied similarly to the electron case except using track P_T instead of E_T.

There is one additional cut to require COT track passing all 96 layers for CMX. It is equivalent to $\rho_{COT} > 140cm$. This variable is defined as

$$\rho_{\mathrm{COT}} = \frac{\mathrm{sign}(\eta) \cdot \mathrm{z_{COT}} - \mathrm{z_0}}{\tan(\lambda)}, \tag{4.2}$$

where $\lambda = \pi/2 - \theta$, $\theta = 2 \cdot \tan^{-1}(e^{\eta})$ and $\mathrm{Z_{COT}} = 155$ cm is used for the length of the COT. This cut is chosen to get good consistency between MC and data when the event is triggered by CMX.

A summary of stubbed muon selection is listed in Table 4.4. The stub reconstruction efficiencies are $1.002 \pm 0.08\%$ and $92.1 \pm 0.7\%$ for CMX and CMUP muons. The

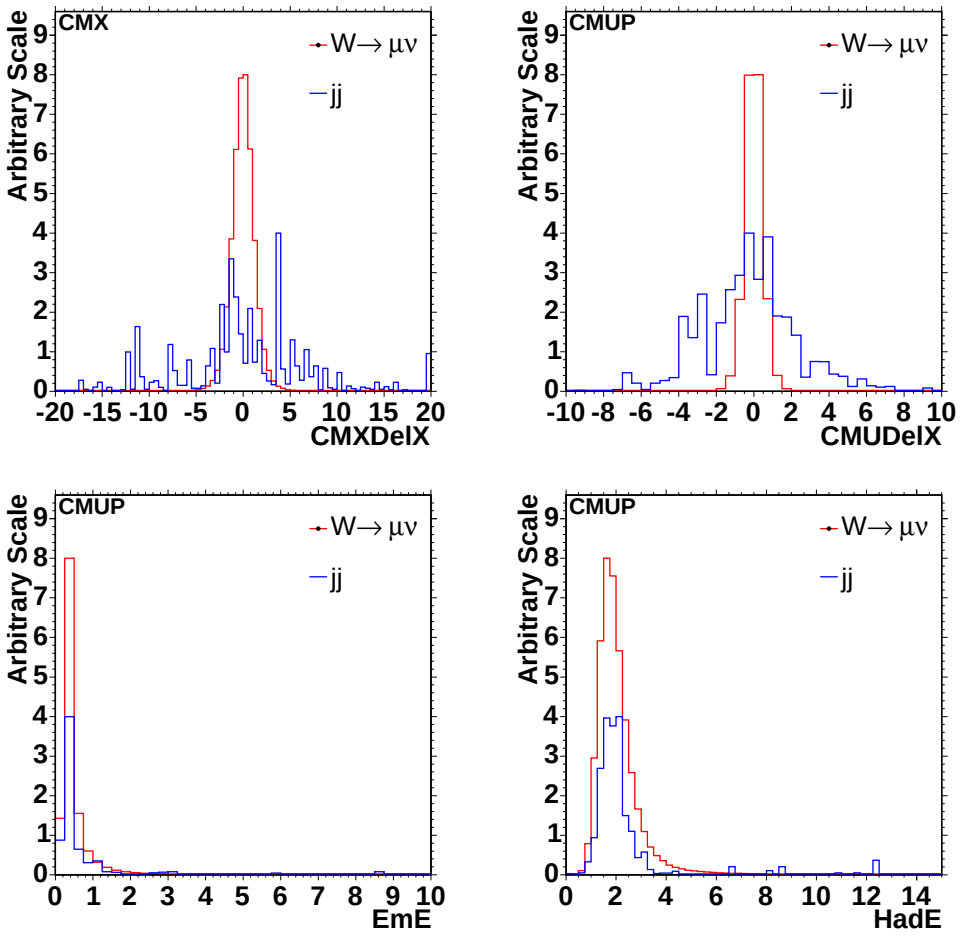


Figure 4.5: The N-1 distribution of the identification variables for CMX/CMUP.

N-1 cut rejection efficiencies are 2 to 3% for EM Energy, Had Energy and Isolation. The identification efficiency for CMX varies from $93 \pm 0.1\%$ to $87.7 \pm 1.7\%$ For CMUP muons, the identification efficiencies are $92.6 \pm 0.9\%$ to $86 \pm 1\%$, respectively. The identification efficiencies without track isolation are $94 \pm 2\%$ to $90 \pm 2\%$ for CMX and $93.9 \pm 0.9\%$ to $89 \pm 1\%$ for CMUP. Figure 4.5 shows the N-1 distributions for identification cuts of muons.

Stubless Central/Forward Muon (CMIOCES/CMIOPES)

A stubless muon is a track passing the minimum ionizing particle criteria, but not matched to a muon stub or matching to a muon stub that failed quality cuts. The

Table 4.5: Lepton identification cuts for CMIOCES and CMIOPES.

	CMIOCES/CMIOPES
E_{em}	$\leq 2 + max(0, (p-100) * 0.0115)$
E_{had}	$\leq 6 + max(0, (p-100) * 0.028)$
Iso/Pt	≤ 0.1
Uniqueness	Not a CMUP or CMX muon
Track $\lvert Z_0 \rvert$	$\leq 60cm$
Track $\lvert D_0 \rvert$	$\leq 0.2cm$ (0.02cm if NSiHit> 0)
$E_{em} + E_{had}$	$> 0.1GeV$
Central	Track CES Fiducial
NAxL(5 hits)	≥ 3
NStL(5 hits)	≥ 3
χ^2/dof	≤ 3
	BcTrk (Larry Correction if Data)
Forward	Track PES Fiducial
Cot Hit Fraction	> 0.6
	No beam constriant on IO tracks

CMIOCES N-1 rejection efficiencies are 4%, 5% and 6% for EM energy, Hadronic energy and Isolation cuts. Each cut gives approximately 3% inefficiencies for CMIOPES. The track for CMIOPES is a combination of COT and SVX hits. An expected number of COT hits is calculated and compared to the real number of COT hits as a quality cut to reject random backgrounds.

The identification efficiency with track isolation for CMIOCES is $82 \pm 1\%$ and it is $82 \pm 1\%$ for CMIOPES. The identification efficiency without track isolation for CMIOCES is $86 \pm 1\%$ and $86 \pm 2\%$ for CMIOPES. Table 4.5 shows the lepton identification selections for stubless muons.

The acceptance increase due to adding CMIOPES is $\sim$5% for $H \rightarrow WW^*$ and $\sim$ 5% to the total backgrounds.

Crack Track

The CrkTrk category includes all tracks not fiducial to the CES nor PES and has the same track selection as CMIOCES with additional conversion veto, but with no calorimeter energy requirements (still calorimeter isolated). The calculation of calorimeter isolation is the isolation energy subtracted from the track P_T to avoid self vetoing the electron type CrkTrk. The N-1 distribution of calorimeter isolation is shown in the Figure 4.6. For electron-type crack track, the N-1 rejection efficiencies are 4%, 6% and 3%

Table 4.6: Lepton identification cuts for CrkTrk.

	CrkTrk
Iso/Pt	≤ 0.1 using CDF Muon *or*
	≤ 0.1 using nearest CDF EMObj with $\Delta R < 0.05$
Track $\lvert Z_0 \rvert$	$\leq 60cm$
Track $\lvert D_0 \rvert$	$\leq 0.2cm$ (0.02cm if NSiHit> 0)
χ^2/dof	≤ 3
NAxL(5 hits)	≥ 3
NStL(5 hits)	≥ 3
Uniqueness	Not a CMUP or CMX muon
Is in Crack	Not Track CES or PES Fiducial
Conversion	$! = 1$
	BcTrk (Larry Correction if Data)

for axial, stereo requirements and conversion requirements, respectively. The isolation rejection efficiency is 8%. For muon-type crack track, 8% and 6% are rejection efficiencies for the number of axial and stereo segments. The isolation and the number of stereo segments have the most impact. The identification efficiency for electron-type CrkTrk is $78 \pm 2\%$. It is $82 \pm 2\%$ without track isolation. The identification efficiency for muon-type CrkTrk is $85 \pm 1\%$. The efficiency without track isolation cut is $87 \pm 1\%$. There's no luminosity dependence observed, however, the fluctuations between runs periods is as much as 5%. Table 4.6 lists the identification selections for CrkTrk.

The acceptance increase due to adding CrkTrk is $\sim$26% for $H \rightarrow WW^*$ and $\sim$ 25% to the total backgrounds.

4.2.2 Jets

Jets are groups of particles formed by the hadronization of quarks and gluons into colorless bound states, known as hadrons. They are reconstructed as multiple tower clusters in both the EM and Hadronic calorimeter. In order to do jet counting, we do not consider the jet is a jet within a cone size of $\Delta\mathrm{R}(\mathrm{jet}, \mathrm{l}) < 0.4$ with respect to an identified lepton. This section reviews the JetClu algorithm adopted for jet reconstructions and jet energy corrections. More details refer to [45].

The JetClu algorithm is started with a list of "seed towers", defined as calorimeter towers with $E_T > 1GeV$. Towers inside a cone of $\Delta R = 0.4$ around the seed tower are merged to form a proto-jet. The center of mass weighting algorithm is applied to the jet with tower centroid position and transverse energy as weights. An iterative pro-

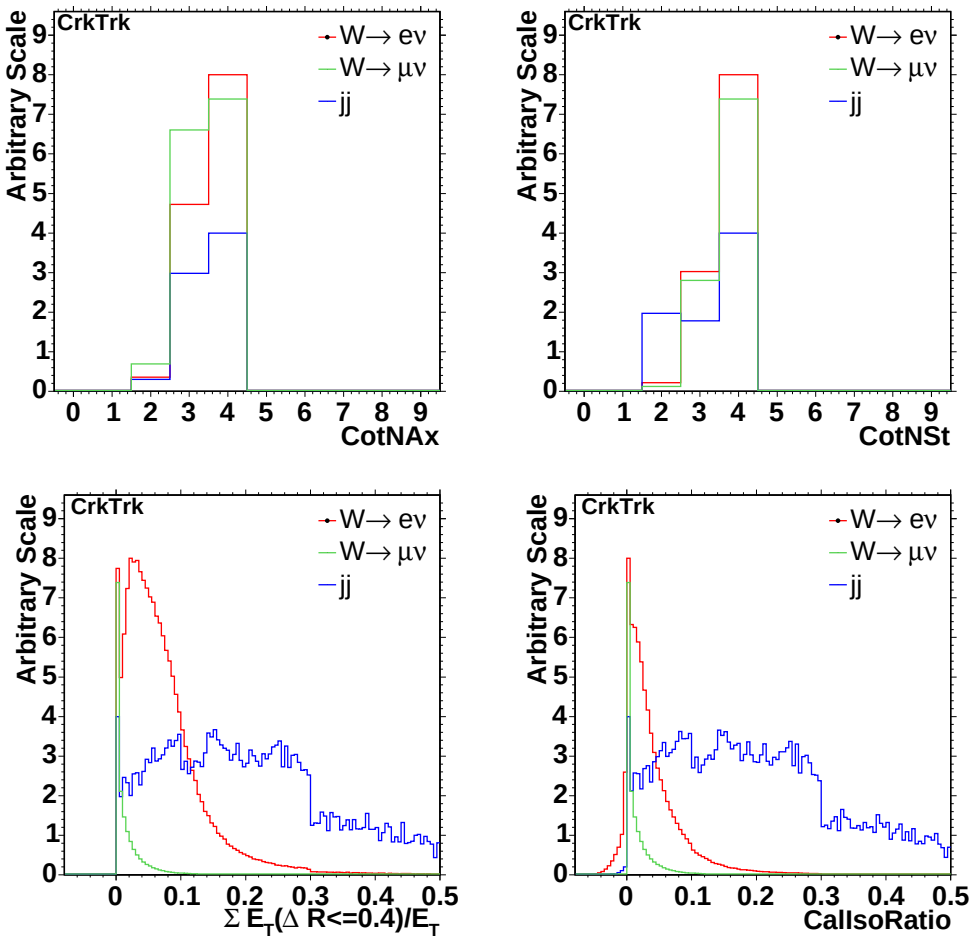


Figure 4.6: The N-1 distribution of the identification variables for CrkTrk.

cedure attempts to split and merge towers from jets to make sure every tower belongs to one and only one jet in each event. A typical central jet found with a cone of radius $\Delta R = 0.4$ has 15 towers.

Since the calorimeter is not 100% hermetic, there is a large amount of energy that is absorbed by non-active material or lost prior to the calorimeter that needs to be accounted for. A series of corrections determined by using the dijet triggered data is applied one by one for total of 5 levels, as follows. Level 1 - Relative corrections: corrects to ensure the jet energy scale is uniform in η and across the different calorimeter systems (CHA, WHA) Level 2 - Stability corrections: corrects for time dependence and aging in the phototubes. Level 3 - Raw energy scaling of central calorimeter: corrects CEM

Run II response to match that of Run I. Level 4 - Multiple interactions: the energy from different $p\bar{p}$ interactions during the same bunch crossing can fall inside the jet cluster, increasing the energy of the measured jet. This is corrected by subtracting this contribution on average, where the average is derived from minimum bias data and is parameterized as a function of the number of vertices in the event, which closely follows the instantaneous luminosity. Level 5 - Absolute energy scale: corrects for any non-linearity in the calorimeter response and energy loss in the un-instrumented regions of each calorimeter. This is done by taking the $\sum P_T$ of the particles within the cone of $\Delta R < 0.4$ around the parton direction which matched the jet direction with $\Delta R < 0.5$.

There are higher level correction:out of cone and underlying event corrections. They are not applied in order to avoid double counting or under counting of the energy in the $\not\!\!E_T$ correction.

4.2.3 Missing Transverse Energy ($\not\!\!E_T$)

The $\not\!\!E_T$ calculation used for this analysis starts with the sum of the transverse energy over the entire calorimeter with respect to the highest $\sum P_T$ vertex as interaction points. The $\not\!\!E_T$ is then corrected for the identified muons by subtracting the muons P_T and adding back its calorimeter E_T. Finally the $\not\!\!E_T$ is corrected for jets $E_T > 15$ GeV and $|\eta| < 2.5$ which do not match any identified leptons within $\Delta R < 0.4$ in the event. The correction here is to subtract the corrected jet E_T and add back the raw E_T. The CrkTrk category of leptons are treated as muons in the $\not\!\!E_T$ calculation. This is necessary because even for electrons in the cracks we do not know how much energy was lost.

4.3 High-level Cuts

This section describes the high-level cuts for the basic candidate selection, a.k.a. Base region, in the fitting. In a simple "cut-and-count" analysis for yield over background, the strategy is to select events with the best anticipated Higgs signal over square root of total yield. However, our strategy is to make a loose selection and let the fitter discriminates Higgs signal from background by using all the available kinematic information. We are optimizing the best expected sensitivity for Higgs observation. The event selections are designed to have sufficient statistics for fitting but remove large background components which would lead to large systematics.

Table 4.7: Event selections for $H \rightarrow WW^*$ analysis (Base region) and ZZ analysis (ZZ region).

High Level Cuts	Vetoed Process
Only two leptons with $P_T >$ 20 and $>$ 10 GeV/c	$WZ, Z\gamma$/jets
Dilepton invariant mass, $m_{ll} > 16$ GeV/c^2	Heavy flavor ($b\bar{b}$, $c\bar{c}$)
Opposite-sign dileptons	$W\gamma$/jets
$\not\!\!E_T^{spec} >$ 25 GeV/c or 15 GeV/c for $e\mu$	Z/γ^*
$N_{jets} < 2$	$t\bar{t}$
Cosmic-ray rejection	Cosmic ray
$\Delta z_0 <$ 4 cm	Decay in fly
ZZ search only	
same flavor	$Z \rightarrow \tau\tau$
$\not\!\!E_T^{sig} > 2.5$[GeV/c]$^{1/2}$	Z+jets

Base region is a set of minimum cuts to suppress major backgrounds: Drell-Yan, $W\gamma$, W+jets and $t\bar{t}$. Each cut is motivated by either satisfying a trigger criteria or rejecting a specific background. Drell-Yan has a large overall cross-section and has no real $\not\!\!E_T$ but enters our high-$\not\!\!E_T$ dilepton sample through mis-measurement tails. The $W\gamma$ and W+jets are selected when a photon conversion is reconstructed as single energetic electron or a hadronized jet identified as a lepton. $t\bar{t}$ could decay to $ll + \not\!\!E_T + 2b$-jets which have the same final state as the signal except for the number of jets. The rest of backgrounds are continuum WW, WZ, and ZZ. We mainly rely on the multivariate technique to discriminate these from signal.

Throughout the remaining sections, the distributions of cut variables and the choice of the cuts are described in detail. All the backgrounds are normalized to the expected yields. The methods to estimate the yields of each processes are explained in chapter 5. The basic selection used to create the sample for WW^* analysis is referred to as the "Base selection" or "Base Region" while it is "ZZ Region" for ZZ search. A brief overview of the high-level cuts are shown in Table 4.7.

Dilepton $P_T > 20$ and > 10 GeV/c Selection

There are two and only two leptons selected in each event to reject three or more lepton events, e.g. $WZ \rightarrow 3l\nu$, $ZZ \rightarrow 4l$ or $Z\gamma$/jets where γ conversions or jets faking leptons. At least one lepton is required to satisfy the trigger criteria and have $P_T > 20$ GeV. We loosen this requirement to 10 GeV for the other lepton to increase

the acceptance. Figure 4.7 shows the generator level leading and sub-leading lepton P_T and η distributions for Higgs, WW and ZZ. Each sample is normalized to the expected yields at 1.9fb^{-1} without any event selections. These plots show that a lower threshold of the second lepton P_T cut is particularly important to increase the acceptance for a lower mass Higgs. However, we didn't go below $P_T < 10$ GeV/c to avoid introducing low P_T leptons decaying from heavy flavor backgrounds. Although there are seven types of leptons, there are only 22 combinations of dilepton categories due to the requirement that one of the leptons has to be a triggerable lepton. The breakdown of these dilepton categories is shown in Table 4.8.

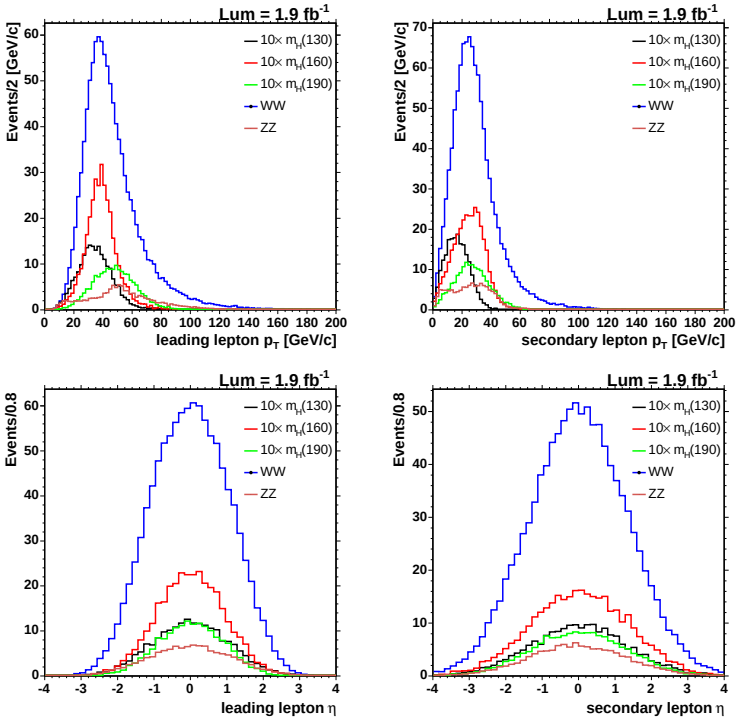


Figure 4.7: The P_T and η distribution of leading and sub-leading letptons (e or μ) without any event selection. Each samples are normalized to expected yields at 1.9fb^{-1}. A factor of 10 is applied to Higgs samples to be compared to WW and ZZ.

Dilepton mass

We require that the invariant mass of the dilepton pair must be greater than 16 GeV/c^2 to avoid introducing heavy flavor backgrounds, e.g. Υ. Figure 4.8(a) shows the predicted dilepton mass distributions after Base selection cuts. The γ^* component of

Table 4.8: At least one lepton must be triggerable and has $P_T > 20$ GeV/c in each dilepton category. There are four out of total seven lepton types are triggerable. The total number of dilepton categories is 22.

Flavor Type	Dilepton Category	Number
ee	**TCE-TCE TCE-PHX PHX-PHX**	3
eμ	**TCE-CMUP TCE-CMX** **TCE**-CMIOCES **TCE**-CMIOPES **PHX-CMUP PHX-CMX** **PHX**-CMIOCES **PHX**-CMIOPES	8
$\mu\mu$	**CMUP-CMUP CMUP-CMX CMX-CMX** **CMUP**-CMIOCES **CMUP**-CMIOPES **CMX**-CMIOCES **CMX**-CMIOPES	7
eTrk	**TCE**-CrkTrk **PHX**-CrkTrk	2
μTrk	**CMUP**-CrkTrk **CMX**-CrkTrk	2
Total		22

DY is growing up exponentially toward low dilepton mass. It is highly suppressed by the $(20, 10$ GeV/c) P_T cuts such that there is a turn-over at 40 GeV/c^2. The cutoff of dilepton mass below 15 GeV/c^2 is due to the Monte Carlo generation threshold of Z mass. The 16 GeV/c^2 event selection cut is chosen slightly higher than 15 GeV/c^2 to remove the mis-calculation of acceptance on the threshold due to detector resolutions. Figure 4.8(b) shows the dilepton mass distributions for Higgs, WW and $W\gamma$ in linear scale. The Higgs acceptance increases if we lower the dilepton mass cut. The improvement of acceptance is larger for low mass Higgs. $W\gamma$ shows similar dilepton mass distributions to Higgs. The lower the Higgs mass is the similar the kinematics are.

$\not\!\!E_T^{spec}$

This a combination of the $\not\!\!E_T$ and $\Delta\phi(\vec{\not\!\!E}_T, \text{lepton, jet})$ cuts, where $\Delta\phi(\not\!\!E_T$, lepton, jet) is the azimuthal angle between $\not\!\!E_T$ and the nearest lepton or jet. The motivation is to have a variable which allows one lepton or jet in a Drell-Yan event to be catastrophically mis-measured without having that event enter the sample. The definition of $\not\!\!E_T^{spec}$ cut is described as below:

$$\not\!\!E_T^{spec} \equiv \begin{cases} \not\!\!E_T & \text{if } \Delta\phi(\vec{\not\!\!E}_T, \text{lepton, jet}) > \frac{\pi}{2} \\ \not\!\!E_T \sin(\Delta\phi(\vec{\not\!\!E}_T, \text{lepton, jet})) & \text{if } \Delta\phi(\vec{\not\!\!E}_T, \text{lepton, jet}) < \frac{\pi}{2} \end{cases}$$

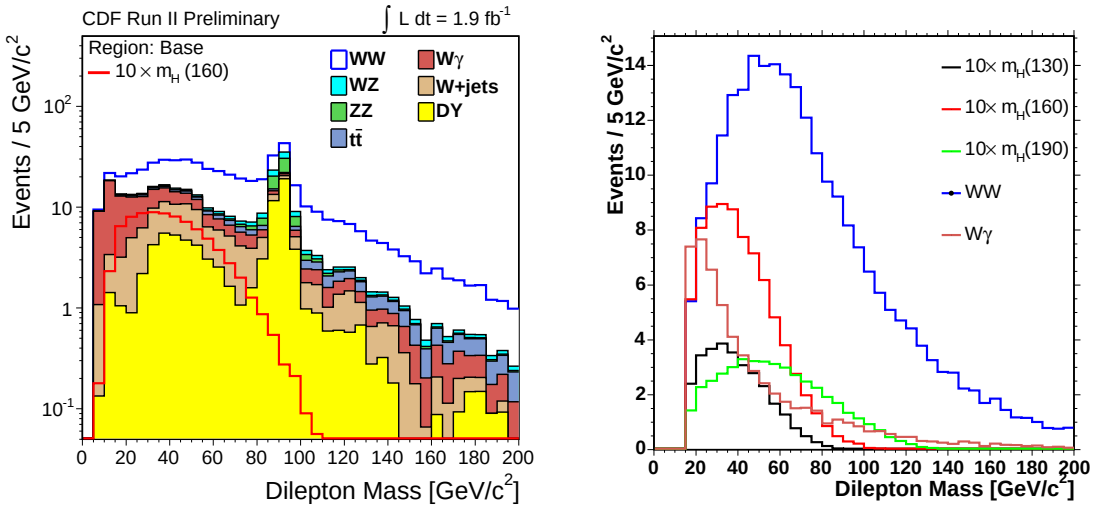


Figure 4.8: The dilepton invariant mass distributions after all the other cuts are applied. The right plot shows the lower the mass threshold the higher the Higgs acceptance is.

where the first case corresponds to gaining energy such that the $\not\!\!E_T$ is pointing away from the lepton or jet, and the second case corresponds to losing energy. In the second case we require that there be missing energy transverse to the lepton or jet. This is demonstrated graphically in Figure 4.9. Figure 4.9 shows the one dimension projections of $\not\!\!E_T^{spec}$ for different dilepton categories. DY backgrounds are largely rejected after requiring high $\not\!\!E_T^{spec}$. To determine the cuts, the optimization of $S/\sqrt{S+B}$ with different cuts are studied where S and B are expected yields for WW and DY. We require $\not\!\!E_T^{spec} > 25$ GeV except $e\mu$ category where we cut on $\not\!\!E_T^{spec} > 15$ GeV. The DY backgrounds in the $e\mu$ category are smaller than other categories by a factor of 10 due to the smaller $e\mu$ decay branching ratio in $Z \to \tau\tau$. The $\not\!\!E_T^{spec}$ is also suppressed more $Z \to \tau\tau$ since boosted τ has collinear decay of e/μ with neutrinos such that $\not\!\!E_T$ is pointing along the lepton direction.

Opposite-sign dileptons

Two leptons must have opposite measured charges. This cut could suppress half of the $W\gamma$ and approximately half of W+jets. Figure 4.11(a) shows the N-1 distribution of charges product of two leptons. More discussions of these two backgrounds are described in Section 5.4.4 and 5.4.6.

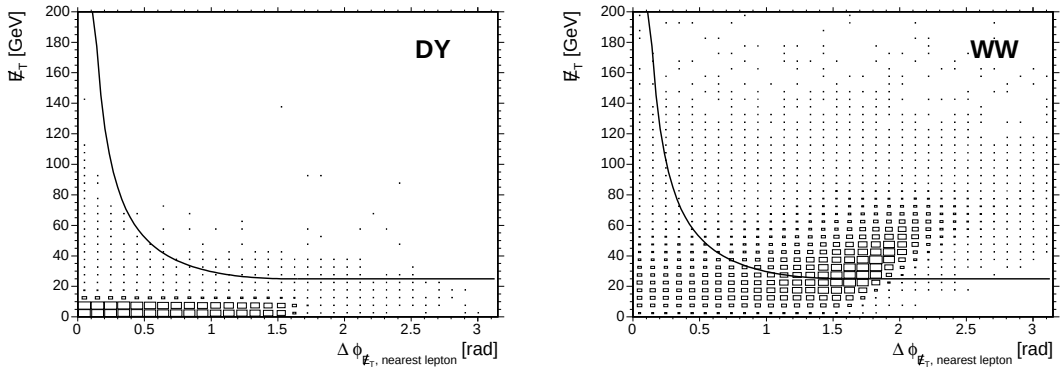


Figure 4.9: Demonstration of the region cut by the $\not\!\!E_T^{spec}$ cut. The solid line is $\not\!\!E_T^{spec} =$ 25GeV/c. This cut effectively reject DY backgrounds with large $\not\!\!E_T$ but small $\Delta\phi$ between $\not\!\!E_T$ and the nearest lepton or jet.

N_{jets}

We require that there be less than two reconstructed jets (Level 5 Jet Energy correction with JetCluCone0.4 algorithm, $JetE_T > 15$ GeV and $|Jet\eta| < 2.5$) in the event to suppress large $t\bar{t}$ backgrounds. Figure 4.11(b) shows the number of jets distribution after requiring all the other event selection cuts.

Cosmic-ray Rejection

To reject cosmic ray events contaminating our muon candidates, we veto events tagged as di-cosmic by the CosmicFinderModule. This is to reject the events with two tracks which are consistent with the assumption of an energetic muon passing through detectors. They are identified by checking the track hit timing information and the impact parameters.

Δz_0

We require that the Δz_0 between two leptons be less than 4 cm because the diboson events we are searching for have decay daughters that originate from a common vertex. This reduces contamination from poorly reconstructed tracks, cosmic rays, multiple interactions, etc.

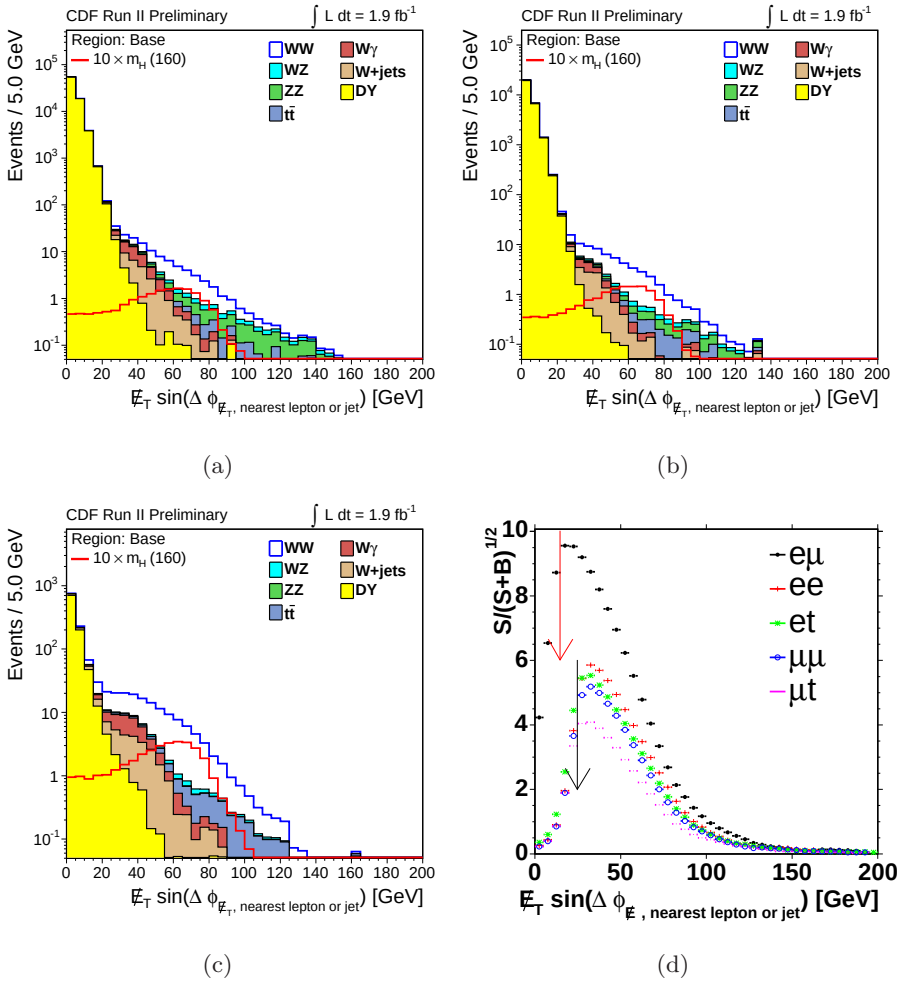


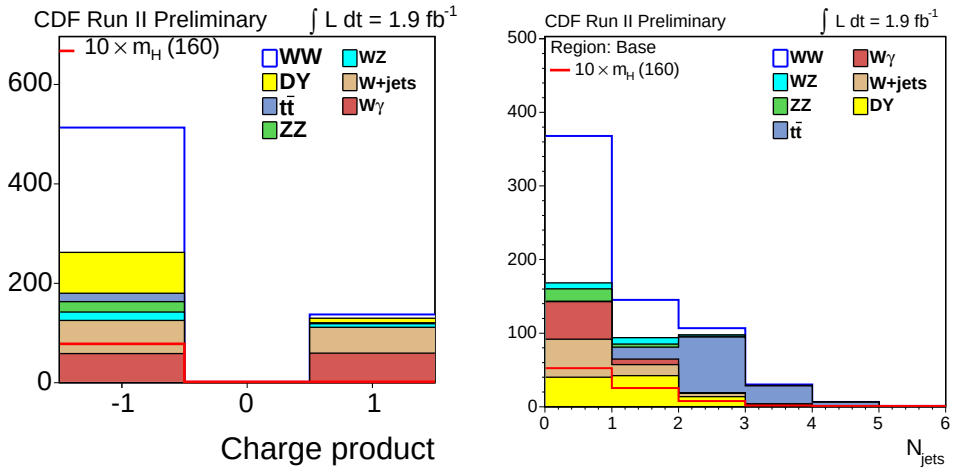
Figure 4.10: The N-1 distributions of $\not\!\!E_T^{spec}$ for (a) ee, (b) $\mu\mu$ and (c) $e\mu$. (d) $S/\sqrt{S+B}$ as a function of $\not\!\!E_T^{spec}$ cut for different dilepton flavors, where S and B are the expected yields of WW and DY at 1.9 fb^{-1}.

Same Flavor

The same flavor requirement is only applied to ZZ search. The $e\mu$ category doesn't add extra sensitivity for ZZ search and introduce more $Z \to \tau\tau$ backgrounds. It also kills 50% of the WW background. For the sake of simplification, only same flavor category: ee, $\mu\mu$, et and μt, are used for ZZ search.

$\not\!\!E_T^{\text{sig}}$

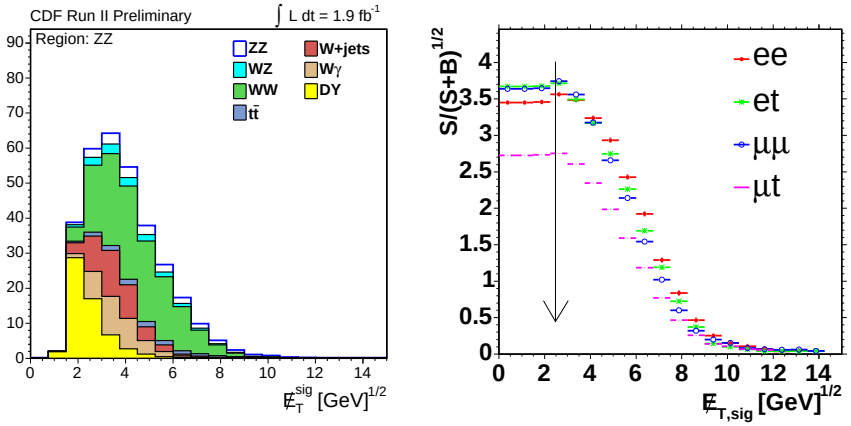
$\not\!\!E_T^{\text{sig}}$ is defined as $\not\!\!E_T/\sqrt{\sum E_T}$, where $\sum E_T$ is the transverse energy sum of all calorimeter towers. The correction of $\sum E_T$ is the same as $\not\!\!E_T$ correction. This variable



(a) (b)

Figure 4.11: The N-1 distributions of (a) two lepton charges product (b) the number of jets.

can suppress events with large un-clustered energy or energetic jet activities. It is only used as a signal discrimination cut for ZZ search to reduce the Z+jets backgrounds. Figure 4.12(a) shows the N-1 distribution in ZZ selection. Figure 4.12(b) shows the choice of the cut with respect to $S/\sqrt{S+B}$ for ZZ analysis. A $\not\!\!E_T{}^{\rm sig} > 2.5$ cut is chosen for ZZ selection. Although this cut is not part of the Base selection, a cross check region is defined to test the un-clustered energy modeling of Monte Carlo.



(a) (b)

Figure 4.12: (a) the N-1 distribution of $\not\!\!E_T{}^{\rm sig}$ distribution. (b) $S/\sqrt{S+B}$ as a function of $\not\!\!E_T{}^{\rm sig}$ cut while S and B are expected yields of ZZ and the other backgrounds. The arrow indicates the event selection cut.

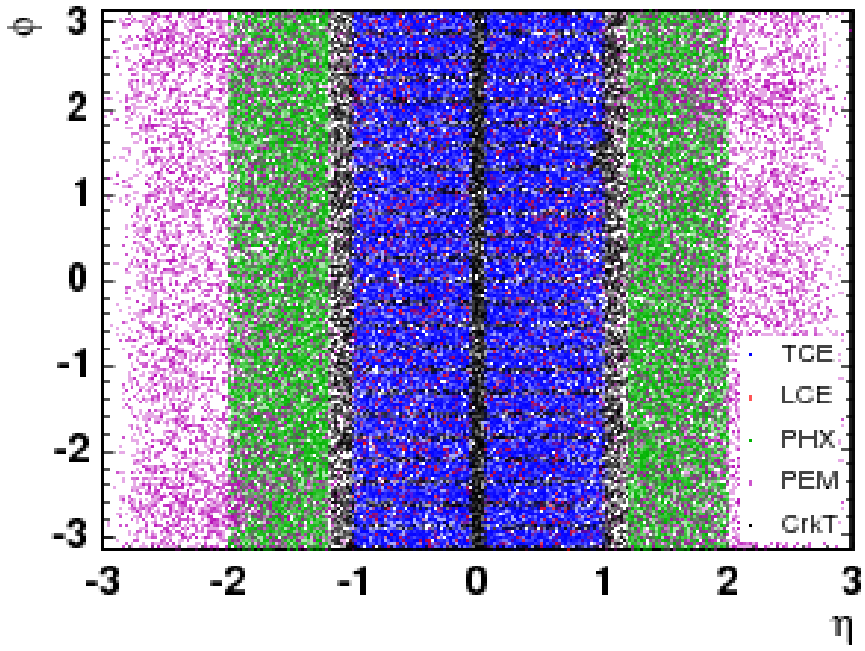
4.4 Acceptance Improvement

Figure 4.13(a) and 4.13(b) shows the η-ϕ coverage of each lepton type. PEM is a forward electron shower without associated tracks. It could be an electron or photon and not used in this analysis. It is used in the WZ analysis while PEM could be constrained by requiring on-shell Z mass cuts [9]. Table 4.9 shows the gains of expected yields by adding two new lepton categories (CMIOPES, CrkTrk). It is calculated from Table 5.18. The overall acceptance gain due to CrkTrk is about 25% and the total acceptance gain is about 40% to 50% for each background. The Higgs signal gain varies from 36% to 49% as m_H varies from 110 GeV/c^2 to 200 GeV/c^2.

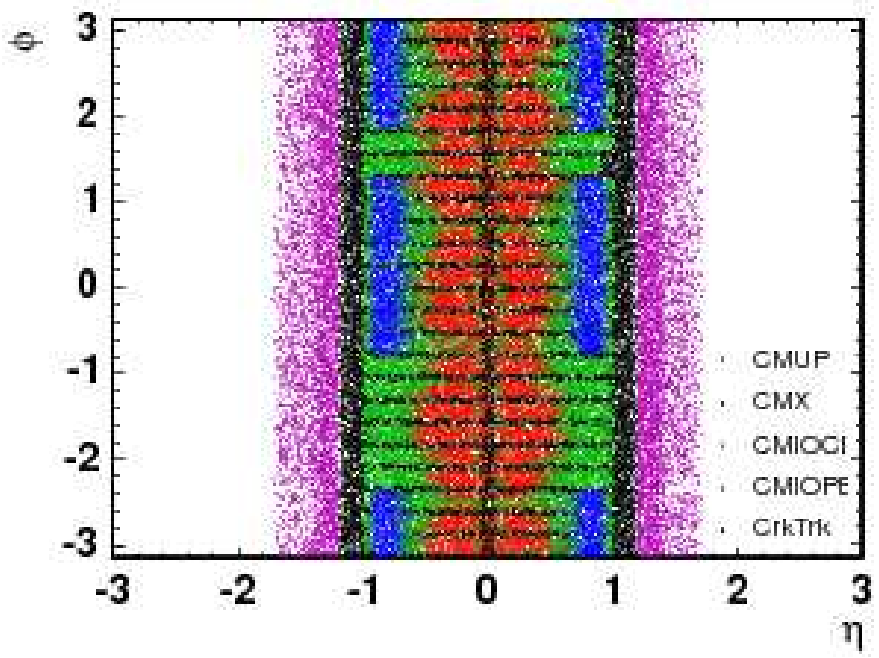
Table 4.9: The expected yields of old and new lepton selections at 1.9 fb^{-1} in Base region. New lepton selection adds two new leptons (CMIOPES, CrkTrk) in addition to old lepton selection (TCE, PHX, CMUP, CMX, CMIOCES).

Category	WW	WZ	ZZ	$t\bar{t}$	DY	$W\gamma$	W+jets	Total
Old Lepton	168.0	11.2	10.4	12.0	57.7	47.8	45.6	352.9 ± 28.6
New Lepton	251.0	16.9	15.0	16.8	82.2	58.5	66.6	507.0 ± 41.1
Gain (%)	49.0	50.8	44.2	40.0	42.5	22.3	46.0	43.7 ± 4.9

	Higgs Mass (GeV/c^2)									
Category	110	120	130	140	150	160	170	180	190	200
Old Lepton	0.3	1.0	2.1	3.3	4.4	5.2	5.1	4.2	3.0	2.4
New Lepton	0.4	1.3	3.0	4.8	6.4	7.8	7.6	6.2	4.4	3.5
Gain (%)	35.6	38.8	41.7	45.9	46.4	48.5	49.5	49.0	48.5	49.0



(a) Monte Carlo $\eta - \phi$ coverage of electron categories used in this thesis are TCE, PHX and CrkTrk. Loose Central Electron (LCE) and Plug EM object (PEM) is only used in the WZ observation [9] which can be constrained by on-shell Z mass cuts.



(b) Monte Carlo $\eta - \phi$ coverage of muon categories used in this thesis. Two new muon types, CMIOPES and CrkTrk, are added to expand the mouon coverage.

Chapter 5

Sample Modeling

The strategy for this analysis is to model every background either from Monte Carlo or data. We first describe the Monte Carlo samples used, followed by an explanation of the data based corrections. We discuss each of the backgrounds in detail. This includes an explanation of the technique for estimating Wjet backgrounds from data. The double counting of Monte Carlo estimated backgrounds in the data driven sample is avoided by matching reconstructed objects to true leptons or photons in Monte Carlo. The final section concludes with a discusssion of the various control regions to cross check the background estimation procedures.

5.1 Monte Carlo Samples

The use of a Monte Carlo simulation is central to this analysis since the signal yield is extracted by performing a hypothesis test on the data for the given predicted model. The analysis results depend on the quality of the simulation of the event generator, detector geometry, and the interactions of particles with the detector material. The signal events and most of the background sample are generated by Pythia 6.216 [37] , Baur [59] and MC@NLO [60] Monte Carlo. Pythia is a multi purpose event generator combining many different hard processes with fragmentation based principally on the string model. The current version is 6.216, this is based on 6.214 with an important bug fix implemented at Fermilab [61]. A detailed simulation of the CDF detector based on the GEANT3 [62] package is used to model the detector's response. Most of the particle interactions with matter are simulated by the standard GEANT algorithms except for a few special cases. For instance, the charged particle ionization and drift properties in

Table 5.1: Higgs Monte Carlo samples used in this analysis. The production cross seciton is calculated to next-to-next-to-leading-logrithm. The $\mathcal{B}(W \to l\nu) = 3 \times 0.108$ [1] has to be applied additionally.

M_HGeV/c^2	Generator	σ_{NNLL} (pb)	$\mathcal{B}(H \to WW)$	Nevts
110	Pythia	1.2808	0.0441	0.5M
120	Pythia	1.0062	0.1320	0.5M
130	Pythia	0.8013	0.2869	0.5M
140	Pythia	0.6455	0.4833	0.5M
150	Pythia	0.5251	0.6817	0.5M
160	Pythia	0.4310	0.9011	1M
170	Pythia	0.3566	0.9653	0.5M
180	Pythia	0.2972	0.9345	0.5M
190	Pythia	0.2493	0.7761	0.5M
200	Pythia	0.2105	0.7347	0.5M

Table 5.2: Monte Carlo samples for background estimation in this analysis.

mode	Generator	$\sigma \times \mathcal{B}$ (pb)	K-factor[a]	Filter Eff	Nevts
$W\gamma$	Baur	13.6	1.36	1.0	9M
WW	MCNLO	1.27	1.0	1.0	4.5M
WZ	Pythia	0.365	1.0	0.76	8.5M
ZZ	Pythia	1.512	1.0	0.23	8.5M
$t\bar{t}$	Pythia	6.7×0.1026	1.0	1.0	9M
$Z(ee, \mu\mu)$ m$_{ll} > 20$GeV/c^2	Pythia	355	1.4	1.0	12M
$Z(ee, \mu\mu)$ m$_{ll} < 20$GeV/c^2	Pythia	920	1.4	0.0156	12M
$Z \to \tau\tau$ m$_{ll} > 10$GeV/c^2	Pythia	1272	1.4	0.00713	12M

[a] If cross-section is calculated to Next-leading-order(NLO), the K-factor is one.

the COT are parameterized and tuned to data. Furthermore, the simulation of showers in the various calorimeters is speed up by the use of the parameterized shower development package GFLASH [63]. The parameters of the electromagnetic and hadronic showers are tuned to data. The detector response is produced in the same format as for the real data. Therefore, the same reconstruction package can be applied to both types of events. The CTEQ5L [34] Parton Density Function (PDF) is used to model the longitudinal momentum distribution of the initial-state partons.

Table 5.1 and 5.2 summarize the signal and background samples used in this thesis. For each sample, we list the corresponding event generator, normalization cross-section and the number of events. The Higgs cross-section has been calcualted to next-to-

next-to-leading order (NNLO) and next-to-next-to-leading logrithm (NNLL) [64]. We used the NNLL cross-section which is generally 10% higher than NNLO value. The background samples are normalized to next-to-leading ordr results calculated by running MCFM program [65]. $t\bar{t}$ is retrieved from reference [66]. The K-factor is defined as the ratio of higher order calculation to leading-order calculations.

5.2 Data Derived Corrections to the Monte Carlo

The processes modeled by Monte Carlo are: Higgs, WW, WZ, ZZ, $W\gamma$, $t\bar{t}$ and Drell-Yan (ee, $\mu\mu$, $\tau\tau$). Each Monte Carlo sample is corrected with respect to the scale factors which accounts for the mis-modeling effects. These effects include lepton identification efficiencies and trigger efficiencies. These efficiencies are measured from $Z-pole$ and compared to Monte Carlo to data to get correction factors. The re-weighted Monte Carlo is therefore an improved description of the real data. The master formula used is:

$$\frac{\sigma \times \mathcal{B} \times \epsilon_{\text{filter}} \times \epsilon_i^{\text{trg}} \times s_i^{\text{lep}} \times \epsilon_{\text{vtx}} \times \mathcal{L}_i}{N_i^{\text{gen}}(|Z_0| < 60 \text{ cm})} \tag{5.1}$$

where

σ	is the cross-section for the Monte Carlo process
$\mathcal{B}$	is any branching fraction for the Monte Carlo process
ϵ_{filter}	is the filter efficiency used in the generation process
ϵ_i^{trg}	is the effective trigger efficiency for the category
s_i^{lep}	is the effective lepton id scale factor
ϵ_{vtx}	is $0.9555 \pm 0.0004(\text{stat}) \pm 0.0031$ (sys)
$\mathcal{L}_i$	is the luminosity of the dataset in which the lepton category falls.

The lepton id scale factors for this analysis are summarized in Table 5.3. They are measured by comparing the lepton identification efficiency in the Monte Carlo and data. The tag-and-probe method is used to measure the identification efficiency in the Z-pole and described in the Appendix B. These are applied by simply multiplying together the scale factor for each lepton type assuming that they are completely uncorrelated.

The trigger efficiencies are 97.7%, 91.0%, 91.9% and 95.9% for TCE, PHX, CMUP and CMX, respectively. They are measured from W or Z resonance from data

Table 5.3: Lepton Id scale factors with and without track isolation cuts. No track isolation cut is applied to the PHX category.

Lepton Category	No Track Iso Cut	with Track Iso < 0.1
0d		
CMUPId	1.014 ± 0.010	1.020 ± 0.011
CMUPReco	0.945 ± 0.007	0.945 ± 0.007
CMXId	1.020 ± 0.013	1.039 ± 0.016
CMXReco	1.002 ± 0.008	1.002 ± 0.008
CMIOCES	1.043 ± 0.016	1.039 ± 0.016
CMIOPES	1.004 ± 0.010	0.991 ± 0.010
CrkTrk μ	0.941 ± 0.013	0.932 ± 0.013
TCE	1.017 ± 0.005	1.009 ± 0.005
PHXTrk	0.999 ± 0.005	
PHXPEM	0.951 ± 0.006	
CrkTrkElec	0.940 ± 0.015	0.924 ± 0.014
0h		
CMUPId	0.992 ± 0.007	0.979 ± 0.009
CMUPReco	0.936 ± 0.006	0.936 ± 0.006
CMXId	0.996 ± 0.011	0.974 ± 0.014
CMXReco	1.005 ± 0.009	1.005 ± 0.009
CMIOCES	1.035 ± 0.012	1.024 ± 0.012
CMIOPES	0.991 ± 0.009	0.968 ± 0.009
CrkTrk μ	0.977 ± 0.009	0.968 ± 0.009
TCE	1.006 ± 0.004	0.995 ± 0.004
PHXTrk	1.010 ± 0.004	
PHXPEM	0.951 ± 0.005	
CrkTrkElec	0.962 ± 0.013	0.933 ± 0.011
0i		
CMUPId	0.979 ± 0.014	0.950 ± 0.018
CMUPReco	0.942 ± 0.012	0.942 ± 0.012
CMXId	1.004 ± 0.020	0.985 ± 0.025
CMXReco	1.015 ± 0.014	1.015 ± 0.014
CMIOCES	1.031 ± 0.023	1.017 ± 0.024
CMIOPES	0.983 ± 0.019	0.969 ± 0.019
CrkTrk μ	0.959 ± 0.017	0.956 ± 0.017
TCE	1.007 ± 0.007	0.979 ± 0.009
PHXTrk	1.013 ± 0.007	
PHXPEM	0.932 ± 0.009	
CrkTrkElec	0.943 ± 0.027	0.941 ± 0.025

and described in Appendix A. These are applied such that if both leptons in the event can trigger we apply the combined efficiency $1-(1-\epsilon_1)(1-\epsilon_1)$. For the case of the MET_PEM trigger we allow either PEM to be part of the trigger, but the MET is only given one chance (this is only important for the PHX-PHX category). In principle, events for which both leptons pass the trigger may be present in more than one primary CDF dataset. We apply a filter such that events like this enter our analysis only once.

Luminosity in Table 4.1 is weighted appropriately to the Monte Carlo samples for each of the lepton pair category.

5.3 Signal

The signal of interest in this analysis is $gg \rightarrow H \rightarrow WW^*$. The fully simulated events are generated by PYTHIA. The production cross-section is normalized to NNLO calculation. The k-factor is 2.1 due to the large soft QCD corrections in top triangle gg fusion diagrams. Although the rate normalization is corrected to NNLO, the p_T and rapidity distribution of Higgs are still the same as PYTHIA default. The effects of the variation of p_T spectrum is estimated through re-weighting method and included as one of the systematics uncertainty. More details see section 7.2.2.

Table 5.4 shows the expected yields and acceptance of $m_H = 160$ GeV/c^2. The acceptance is calculated with respect to the total number of $H \rightarrow WW$ events. Table 5.5 shows the expected yields as a function of Higgs mass. Figure 5.1 shows the acceptance as a function of Higgs mass and also compared to WW and ZZ.

Table 5.4: The expected yields for of Higgs $m_H = 160$ GeV/c^2 in the base region. The acceptance is calculated with respect to the total numebr of $H \rightarrow WW$ events.

Cut	ee	$e\mu$	$\mu\mu$	et	$\mu\ t$	$ll(\pm stat)$
TrigDil	2.1	4.1	1.8	1.7	1.1	10.9 ± 0.1
maxΔz	2.1	4.1	1.8	1.7	1.1	10.9 ± 0.1
m_{ll}	2.0	4.0	1.8	1.7	1.0	10.5 ± 0.1
$\not\!\!E_T^{spec}$	1.6	3.6	1.5	1.4	0.9	8.9 ± 0.1
OS	1.6	3.5	1.5	1.4	0.9	8.7 ± 0.1
Njets<2	1.4	3.1	1.3	1.2	0.8	7.8 ± 0.1
0jet	1.0	2.0	0.9	0.8	0.5	5.2 ± 0.1
A(Njets<2)	1.9e-03	4.2e-03	1.8e-03	1.7e-03	1.0e-03	$1.056e-02 \pm 2e-04$

Table 5.5: Predicted Standard Model Higgs yields in the base region as a function of Higgs mass.

	Higgs Mass (GeV)									
Category	110	120	130	140	150	160	170	180	190	200
$e\ e$	0.1	0.3	0.6	0.9	1.2	1.4	1.4	1.1	0.8	0.6
$e\ \mu$	0.2	0.6	1.3	2.0	2.6	3.1	3.0	2.5	1.8	1.4
$\mu\ \mu$	0.1	0.2	0.5	0.8	1.1	1.3	1.3	1.0	0.7	0.6
e trk	0.0	0.2	0.4	0.7	0.9	1.2	1.2	1.0	0.7	0.6
μ trk	0.0	0.1	0.2	0.4	0.6	0.8	0.7	0.6	0.4	0.3
Total	0.4	1.3	3.0	4.8	6.4	7.8	7.6	6.2	4.4	3.5

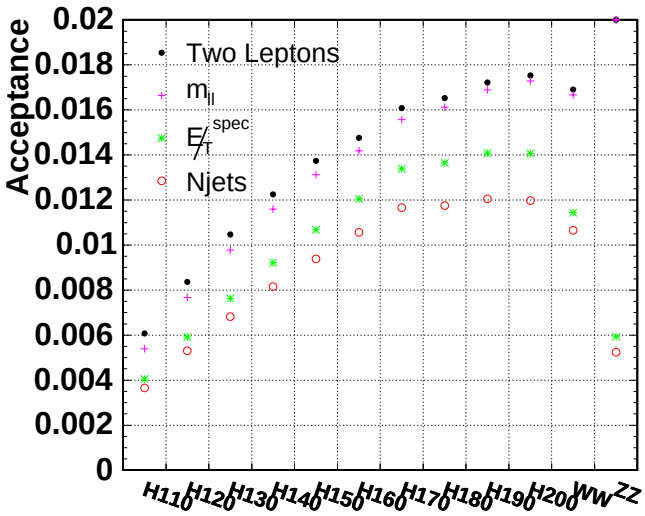


Figure 5.1: The Higgs acceptance with four consecutive event selection cuts. The denominator is $H \rightarrow WW$ process. WW and ZZ are both shown in the plot. PYTHIA ZZ which decays generically and pass two lepton (e or μ) filter cuts is used as the denominator.

5.4 Backgrounds

Backgrounds to dilepton plus $\not\!\!E_T$ production are dominated by mis-reconstructed leptons or $\not\!\!E_T$ and additional background from other physics processes that overlap in phase space and final states. This section describes each of these backgrounds, why they contaminate the signal, how to suppress and estimate them.

Table 5.6: The expected yields for DY in the Base region.

Cut	ee	$e\mu$	$\mu\mu$	et	$\mu\ t$	$ll(\pm stat)$
TrigDil	85467.1	1110.2	65182.2	29637.1	21622.1	203018.8 ± 98.6
maxΔz	85433.7	1109.2	65122.2	29629.5	21618.4	202913.1 ± 98.6
m_{ll}	85341.4	1107.5	65033.0	29603.2	21594.7	202679.8 ± 98.5
$\not\!\!E_T^{spec}$	32.6	36.0	20.3	13.2	6.6	108.6 ± 2.0
OS	30.4	30.2	20.2	12.0	6.3	99.1 ± 1.9
Njets<2	26.6	22.5	17.6	10.3	5.3	82.2 ± 1.8
0jet	16.4	5.7	9.3	6.0	2.6	40.0 ± 1.4
A(Njets<2)	2.6e-06	2.2e-06	1.7e-06	1.0e-06	5.2e-07	$8.096e-06 \pm 2e-07$

5.4.1 Drell-Yan

One of the major sources of this background is the large fake $\not\!\!E_T$ due to mis-measurement of lepton's energy or jets energy. In this case, the $\not\!\!E_T$ direction will be collinear with the most mis-measured lepton's or jet's direction. A $\not\!\!E_T^{spec} > 25$GeV cut reduces significantly the amount of DY in the order of O(10^{-4}) in the e^+e^- and $\mu^+\mu^-$ categories. The $e\mu$ category actually contains real $\not\!\!E_T$ contribution from $DY \to \tau\tau$. The $\not\!\!E_T^{spec} > 15$ GeV cut reduction is about O(10^{-3}). In the case of Z+jets, the fake $\not\!\!E_T$ could be coming from missing the jets in the reconstruction. If Z gains large transverse energy due to the recoil from an energetic jet, it could mimic a Higgs signal since the azimuthal angle of two leptons are boosted to be smaller.

Another source of fake $\not\!\!E_T$ is coming from the uncertainty of reconstructing un-clustered energy. The un-clustered energy originates from the soft jets of underlying events, pile-up events, beam remnants or PMT response. The larger the total energy deposit in the detector, the larger the fluctuation of $\not\!\!E_T$ magnitude. The fake $\not\!\!E_T$ due to un-clustered energy could be suppressed by $\not\!\!E_T{}^{\rm sig}$. It's used as a variable to define cross check region but not part of the base region criteria. However, to further suppress DY backgrounds in ZZ analysis, a $\not\!\!E_T{}^{\rm sig} > 2.5$ cut is applied in the ZZ search.

The estimate of this background is from 36M DY Monte Carlo samples. The ISR/FSR simulation is included, which covers the contribution from radiative photons and QCD partons. The predicted yield of DY in the base region is 20.9 events. Table 5.6 shows the acceptance of DY normalized to the Drell-Yan production with $m_{ll} > 10$GeV/c^2

Figure 5.2(a) shows the reconstructed dilepton p_T v.s. generated dilepton p_T

for events that have $\not\!\!E_T > 25$ GeV but failed $\not\!\!E_T{}^{spec}$ cut. A large fraction of events have mis-reconstructed lepton energy. Figure 5.2(c) shows a two dimentional distribution for each dilepton candidate and x-axis is the maximum p_T difference of the two leptons. The mis-measured leptons show up in the $\eta = 1.1$ and $\eta = 0$ crack. Figure 5.2(b) shows less number of events of fake $\not\!\!E_T$ in $Z \to \mu\mu$ due to mis-measurement of lepton energy. Figure 5.2(d) shows that the mis-measurement of μ is mostly coming from forward track. Figure 5.2(e) is the $Z \to ee$ events with 1jet where $\not\!\!E_T > 25$ GeV and $\not\!\!E_T^{spec} < 25$ GeV. It shows two leptons are mostly measured well and indicates the large $\not\!\!E_T$ is due to the mis-measurement of jet energy. Figure 5.2(f) shows the $\not\!\!E_T^{spec}$ suppressed events in $Z \to \tau\tau$. These events have true $\not\!\!E_T$ because of the neutrinos from τ decay. Mismeasurement of leptons is thus not the dominant effect.

5.4.2 Heavy Diboson WW

Table 5.7: The expected yields for WW in the Base region.

Cut	ee	$e\mu$	$\mu\mu$	et	$\mu\ t$	$ll(\pm stat)$
TrigDil	81.0	154.9	61.4	65.5	35.6	398.4 ± 1.0
maxΔz	81.0	154.8	61.4	65.4	35.6	398.2 ± 1.0
m_{ll}	80.0	152.7	60.3	64.6	35.1	392.8 ± 1.0
$\not\!\!E_T^{spec}$	51.4	118.4	37.5	40.6	21.5	269.5 ± 0.9
OS	48.4	114.8	37.5	39.3	21.5	261.6 ± 0.8
Njets< 2	46.6	110.1	36.0	37.8	20.6	251.0 ± 0.8
0jet	37.0	87.5	28.6	30.1	16.3	199.5 ± 0.7
A(Njets< 2)	2.0e-03	4.7e-03	1.5e-03	1.6e-03	8.7e-04	$1.065e-02 \pm 3e-05$

WW events decaying leptonically have the same final state as the signal. After $\not\!\!E_T^{spec}$ cuts, it becomes the dominant background for the $H \to WW$ search. The estimation of this background is coming from MC@NLO generator combined with Pythia parton showers and CDF detector simulation. The systematics of NLO acceptance is determined by comparing MC@NLO to Pythia. More details will be discussed in section 7.2.1. The estimated yield of WW in the base region is 251 ± 0.8 (stat) events.

5.4.3 Heavy Diboson WZ, ZZ

WZ events contribute as a dilepton background when both bosons decay leptonically and one of the leptons is not reconstructed or W decays hadronically and Z decays leptonically with mis-measured jets generating large $\not\!\!E_T$. The two and only two

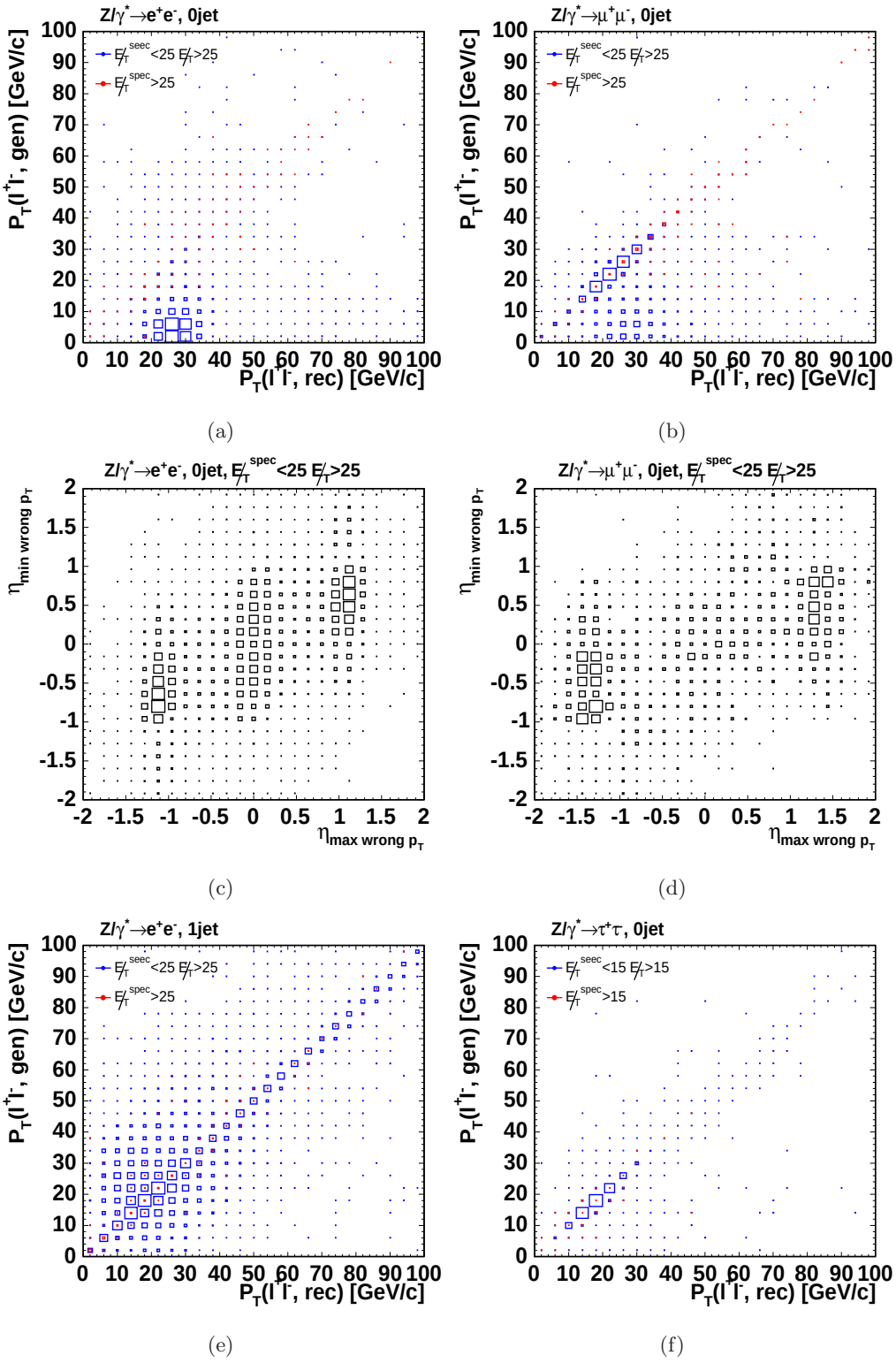
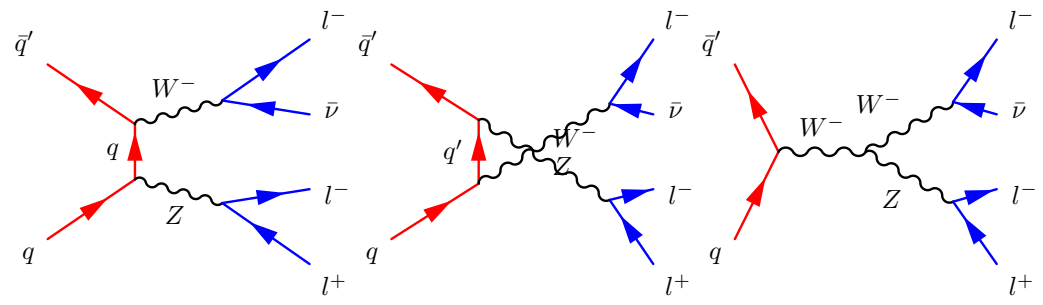


Figure 5.2: The distribution of generator-level dilepton p_T with respect to $\not{E}_T$.



(a) t-channel (b) u-channel (c) s-channel

Figure 5.3: Leading-order diagrams contributing to WZ production at the Tevatron and fractions decaying to final states involving leptons and jets.

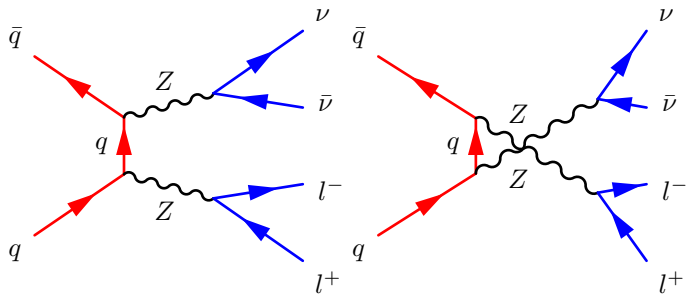
lepton requirement eliminates 50% of the trilepton events and the jet veto suppresses most of the hadronic W final state. The dominant contribution after all the cuts is $WZ \rightarrow l\nu l^+l^-$ with one of the leptons not reconstructed or failing the identification cuts. Figure 5.3 is the leading order diagram for WZ production decaying fully leptonically. The WZ Monte Carlo generated by PYTHIA does not have γ^* component. The resulting expected yields of WZ is 16.9 events. Table 5.8 shows the expected yields and acceptance for WZ backgrounds.

Table 5.8: The expected yields for WZ in the Base region.

Cut	ee	$e\mu$	$\mu\mu$	et	$\mu\ t$	$ll(\pm stat)$
TrigDil	43.2	10.1	32.5	18.1	12.0	115.8 ± 0.1
maxΔz	43.2	10.1	32.4	18.1	12.0	115.8 ± 0.1
m_{ll}	43.1	10.0	32.4	18.1	12.0	115.5 ± 0.1
$\not\!\!E_T^{spec}$	7.8	7.2	5.7	4.2	2.4	27.3 ± 0.0
OS	6.0	3.7	4.6	3.0	1.8	19.1 ± 0.0
Njets<2	5.3	3.2	4.1	2.6	1.6	16.9 ± 0.0
0jet	2.3	1.6	2.3	1.2	0.9	8.2 ± 0.0
A(Njets<2)	7.7e-04	4.6e-04	5.9e-04	3.8e-04	2.3e-04	$2.430e-03 \pm 5e-06$

ZZ production can also end up with two lepton plus large $\not\!\!E_T$ final state. Among those candidates which pass the Base selection criteria, there are 95% coming from $ZZ \rightarrow l^+l^-\nu\bar{\nu}$ and 5% coming from $ZZ \rightarrow 4l$. $ZZ \rightarrow l^+l^-\nu\bar{\nu}$ has the same final state as $H \rightarrow WW^*$ signal but $\not\!\!E_T$ is harder due to the fact that both neutrinos can only decay from Z pole. It is the physics that neutrino only couple to neutral weak boson Z but other charged fermions can have Z/γ^* interference couplings. In the case

of $ZZ \rightarrow 4l$, there are two leptons either not reconstructed or failed the identification cuts such that they cause a large fake $\not\!\!E_T$.



(a) t-channel (b) u-channel

Figure 5.4: Leading order t-channel and u-channel diagrams contribute to ZZ production at the Tevatron.

Figure 5.4 shows the leading order diagram for this process. There are t/u-channel diagrams in the standard model; while s-channel is forbidden. If an excess of production consistent with s-channel prediction is observed, this is new physics. There is one type of s-channel production when ZZ decays to charged fermions. It's essentially a production of a single Z decays to two charged fermions with final state radiation of a virtual γ converts to two charged fermions. This diagram is not included in the Pythia Monte Carlo. We estimated the contribution of this background by computing its cross-section in MadGraph with the $ZZ \rightarrow 4l$ final state and find it is negligible for this analysis.

The $Z/\gamma^*Z/\gamma^*$ is simulated in Pythia. The estimate of ZZ from Pythia in the base region is 15.0 events. Table 5.9 shows the expected yields and acceptance for ZZ production in the Base region.

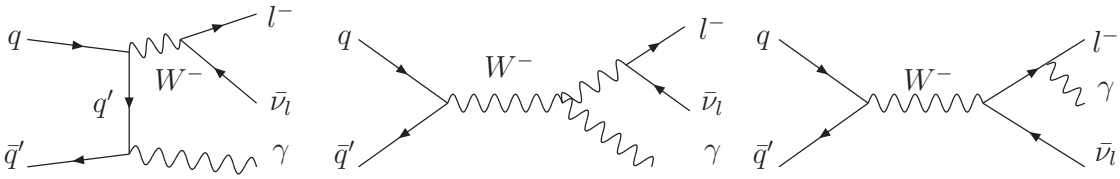
5.4.4 $W\gamma$

$W\gamma$ contributes to dilepton background when the W decays leptonically and the γ converts to an asymmetric e^+e^- pair and reconstructs as an electron. A central photon conversion is an interaction with material before the COT volume, and a plug conversion could be just a plug photon with fake Pheonix track due to silicon noise hits. The estimation of $W\gamma$ is purely from Monte Carlo and an extra 1.3 scale factor is applied from the conversion efficiency study and described in more details in section 7.4. Figure 5.5 shows the leading order diagram for $W\gamma$ production. The collinear divergence

Table 5.9: The expected yields for ZZ in the Base region.

Cut	ee	$e\mu$	$\mu\mu$	et	$\mu\ t$	$ll(\pm stat)$
TrigDil	39.16	1.97	31.66	15.14	10.85	98.79 ± 0.12
maxΔz	39.15	1.97	31.64	15.14	10.85	98.74 ± 0.12
m_{ll}	39.07	1.95	31.58	15.10	10.83	98.54 ± 0.12
$\not{E}_T^{spec}$	6.65	0.79	5.18	2.63	1.80	17.04 ± 0.05
OS	6.20	0.49	5.06	2.49	1.74	15.97 ± 0.05
Njets<2	5.88	0.35	4.80	2.35	1.65	15.03 ± 0.05
0jet	4.74	0.15	3.88	1.87	1.32	11.95 ± 0.04
A(Njets<2)	2.0e-03	1.2e-04	1.7e-03	8.2e-04	5.7e-04	$5.233e-03 \pm 2e-05$

of final state radiation is the dominant production of $W\gamma$. It also plays a significant role to mimic Higgs signal signature since both the azimuthal angle and invariant mass of two leptons are small. The expected yields of $W\gamma$ background is 58.5 events and shown in table 5.10.



(a) t-channel (b) s-channel (c) FSR

Figure 5.5: The leading order Feynman diagram for $W\gamma$ production.

Table 5.10: The expected yields for $W\gamma$ in the Base region.

Cut	ee	$e\mu$	$\mu\mu$	et	$\mu\ t$	ll
TrigDil	116.4	82.4	0.0	29.6	5.2	233.6 ± 1.2
maxΔz	116.2	82.1	0.0	29.5	5.2	233.0 ± 1.2
m_{ll}	86.5	65.0	0.0	22.5	3.6	177.6 ± 1.0
$\not{E}_T^{spec}$	54.9	50.0	0.0	13.1	2.3	120.3 ± 0.9
OS	27.7	24.3	0.0	6.6	1.2	59.7 ± 0.6
Njets<2	27.2	23.8	0.0	6.5	1.1	58.5 ± 0.6
0jet	23.9	20.6	0.0	5.7	1.0	51.2 ± 0.6
A(Njets<2)	1.1e-04	9.6e-05	0.0e+00	2.6e-05	4.6e-06	$2.359e-04 \pm 2e-06$

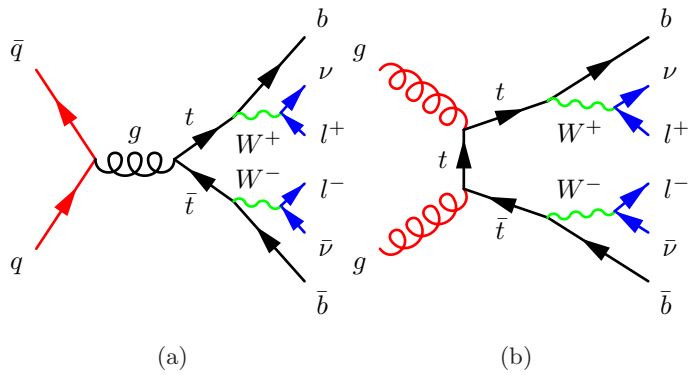


Figure 5.6: The dominant leading order diagrams of $t\bar{t}$ production at Tevatron. The $q\bar{q}$ annihilation s-channel diagram contributes 80% of the total production cross-section.

5.4.5 $t\bar{t}$

Production of $t\bar{t}$ pairs in the dilepton channel $t\bar{t} \rightarrow l\nu l\bar{\nu}b\bar{b}$ results in a final state identical to a $H \rightarrow WW$ candidate apart from the presence of additional b-jets. Figure 5.6 is the dominant leading order diagrams for this background. The $q\bar{q}$ annihilation s-channel diagram contributes 80% of the total production cross-section. The gluon-gluon fusion via t-channel contributes 20% of production rate because of the small PDF values in high-x at $\hat{s}$ =1.96 TeV. There also exists s-channel diagrom in the gluon-gluon fusion process. However, it has s-channel suppression compared to t-channel. Top quark decays 100% to Wb such that it has the same WW final states with additional two b jets. The jet veto cut ($N_{jet} < 2$) can reduce this background by factor of 14%. The remaining background is due to the b-jets being produced in un-instrumented region or the mis-reconstruction of jet energy so that the jet fails the $E_T > 15$GeV cut. The estimation is scaled from PYTHIA Monte Carlo and the predicted yield is 16.8 ± 0.1 events and shown in Table 5.11.

5.4.6 W+jets Background

The dilepton plus $\not\!\!E_T$ signal has a background from the W+jet(s) process when one of the jets is reconstructed as an identified lepton. The leading order Feynman diagram for this type of background is shown in Figure 5.7. The background is suppressed by the isolation requirements on the lepton selection and the opposite sign cut. A jet is usually fragmented into multiple hadrons distributed within a finite cone surrounding the identified lepton. This mis-identified lepton is therefore effectively rejected by requiring

Table 5.11: The expected yields for $t\bar{t}$ in the Base region.

Cut	ee	$e\mu$	$\mu\mu$	et	$\mu\ t$	$ll(\pm stat)$
TrigDil	38.2	76.8	33.3	33.1	19.8	201.2 ± 0.4
maxΔz	38.2	76.8	33.3	33.1	19.8	201.1 ± 0.4
m_{ll}	37.9	76.1	33.0	32.9	19.6	199.4 ± 0.4
$\not\!\!E_T^{spec}$	22.0	54.8	19.2	18.9	11.2	126.1 ± 0.3
OS	21.1	53.6	19.2	18.4	11.2	123.4 ± 0.3
Njets<2	2.9	7.0	2.7	2.6	1.5	16.8 ± 0.1
0jet	0.1	0.3	0.1	0.1	0.1	0.6 ± 0.0
A(Njets<2)	2.3e-04	5.5e-04	2.1e-04	2.1e-04	1.2e-04	$1.322e-03 \pm 9e-06$

isolation cut. The opposite sign requirement reduces the fake dilepton pairs from light quark jets by $\approx 50\%$. There are some charge correlations between the W and the jet which can make this not exactly 50%. A light quark jet generates fake lepton with equal sign probabilities. The ratio of opposite sign to same sign candidates is more complicated in the heavy flavor jets due to mixing. It is therefore a good control region to compare data and MC for same sign and opposite sign candidate to test our sample modeling. To have an idea how a fake lepton originates from jet, a Monte Carlo study is performed in the PYTHIA dijet sample. However, the estimate of W+jets backgrounds is purely data-driven. The lepton fake probabilities ("rates") are measured in the inclusive jet sample after corrections are made for the small contributions from real isolated leptons from electroweak processes. The fake rate is applied to a triggered lepton+fake denominator sample to estimate W+jet yields.

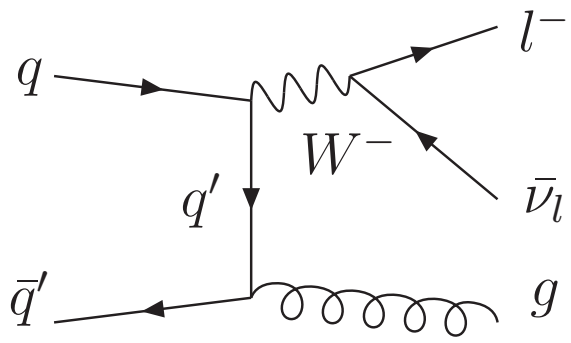
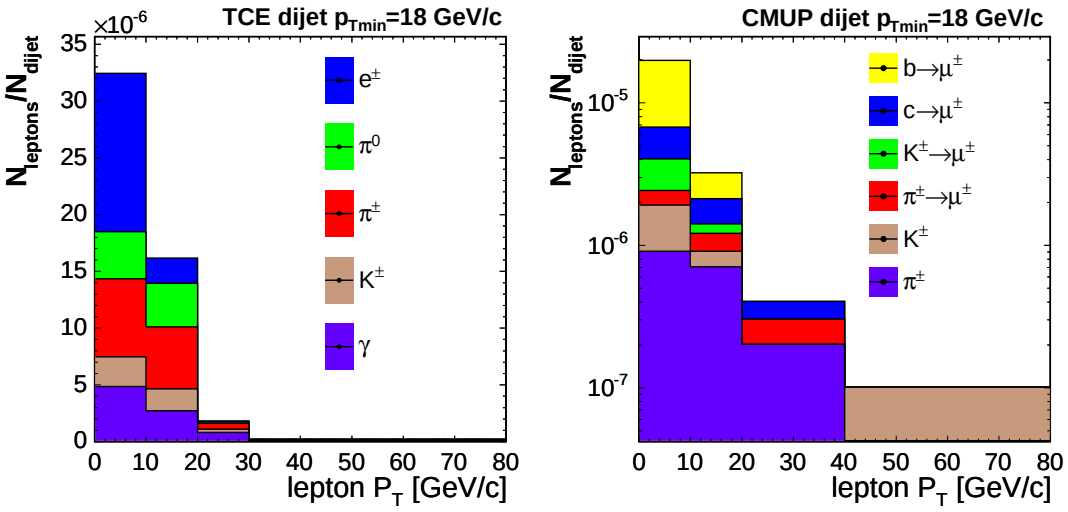


Figure 5.7: Leading Order Feynman diagram for W+jet

The Compositions of Fake Leptons in Dijet Monte Carlo

A fake lepton could be a real lepton decayed from a heavy quark (s, c, b), mis-reconstructed lepton from a hadron ($\pi^{\pm,0}$,$K^{\pm}$,$p^{\pm}$,n) or a conversion electron from a FSR photons radiated from partons. To understand the origin of fakes, we process PYTHIA dijet Monte Carlo and study the events which produce leptons passing the identification criteria. The normalization is defined as the number of reconstructed leptons divided by the total number of dijet events. These results are not used as fake background estimate, but only provide an insight on the origin of fakes. The true estimate is mainly driven from data, and described in the next section.

The analysis method is to do a $\Delta R < 0.1$ matching between an identified lepton to a final state particle (StatusCode=1) in HEPG bank. HEPG bank is the data block that records the generator information of hard interaction particles and parton shower particles generated by PYTHIA. After finding the final state particle, a backward tracing is done to search for fragmented partons (StatusCode=2) and the hard interaction partons (StatusCode=1).



(a) (b)

Figure 5.8: Fake rates of TCE and CMUP as functions of lepton p_T in the dijet PYTHIA monte carlo. The fake leptons can be real leptons originated from semileptonic heavy flavor decay, hadrons or photon conversions. or hadrons.

Figure 5.8(a) shows the p_T distribution of TCEs reconstructed from dijet PYTHIA sample where the jet p_{Tmin} is greater than 18 GeV/c. The dominant sources of fake cen-

tral electrons vary as a function of p_T. The major compositions of fake TCEs in the p_T range 10 to 20 GeV/c are 30%, 25%, 20%, 12% and 13% for $\pi^{\pm}$, γ conversions, π^0/η, $K^{\pm}$ and real electrons, respectively. $\pi^{\pm}$ fakes are mainly through charge exchange reaction, $\pi^- p^+ \rightarrow n\pi^0$. The charged pion produce tracks and neutral pion decaying to photon pairs and deposit energy in the calorimeter. γ fakes are caused by photon conversions and the photons are emitted either via initial state radiation (ISR) from initial state partons or through final state radiation (FSR) from out-going fermions. Neutral pions and ηs both decay to photon pairs and deposit energy in the calorimeter. If one of the photon converts to electron pairs or hadrons overlap with charged pions, they could be mis-reconstructed as TCEs. Kaon fakes electrons either through strong interactions with calorimeter or from a higher resonance decay chain, $K^* \rightarrow K^+\pi^0$, where γ deposit EM energy. The real electrons originate from heavy flavor parton, c and b. Combining with kaon contributions, the total heavy flavor contribution is 25%. The compositions of fake TCE sources in the $p_T > 20$GeV/c range are 50% and 40% for photon conversions and charge exchange reactions.

The p_T distribution of fake CMUP is shown in Figure 5.8(b). The contributions from real muons, charged pions and kaons in p_T range 10 to 20 GeV/c are 50%, 17% and 33%, respectively. Real muons are decayed from heavy flavor b or c. The charged pions could punch through hadron calorimeters and leave a muon stub. The high p_T charged kaon which decayed after passing through COT could provide high p_T tracks and a muon stub. If a kaon decays to a muon in the COT, there may be a kink in the track trajectory which may cause it to be reconstructed as a high p_T track.

In summary, the fake electrons in dijet Monte Carlo are mainly from light flavor jets for lepton $p_T > 10$GeV/c. Although heavy flavor jets can produce real electrons, they tend to have a softer p_T (less than 10 GeV/c) spectrum. The fake muons are mainly coming from heavy flavor jets (b, c, s). Note that, there are more than 60% of the heavy flavor jets branching out from gluon splitting. A rough estimate of the rates for a jet to fake a TCE and CMUP is O(10^{-5}) and O(10^{-6}), respectively. The probability for both jets faking leptons and creating a lepton pair is extremely small O(10^{-10}).

Fake Rate Measurement

The fake rate is defined as the probability of a denominator object to be identified as a real lepton. The denominator object, or referred to as fake-able object, is a

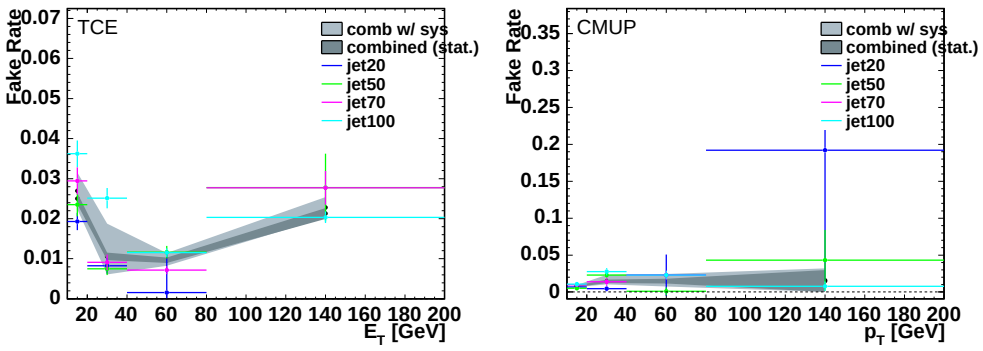


Figure 5.9: Fake rates for TCE and CMUP as a function of p_T. The results are combined from four different dijet triggered data and the gray band is the uncertainty for systematic variation.

loosely selected lepton. The selection is sufficiently loosened that the denominator samples are entirely dominated by jets. The fake rates are obtained from large samples of jet triggered data with E_T thresholds of 20, 50, 70 and 100 GeV. The denominator object matching to the trigger jet in each event is excluded from the fake ratio calculation to reduce the trigger bias. The small contribution of real leptons from W and Z decay are also subtracted from calculation. More details are described in appendix C. Figure 5.9 shows the fake rates for TCE and CMUP as a function of p_T. The combined fake rates from four jet samples with systematics for each lepton are used as the nominal fake rates. The fake rates for TCE and CMUP are 1% to 2% level and the uncertainties are roughly 20%. The definition of fake-able objects and the fake rates of all leptons are described in Appendix C.

Fake Background Estimate

Fake rates are applied to a sample of lepton plus denominator object events where the denominator object is treated exactly as an identified lepton would be. The lepton+denominator pair is rejected if the denominator can pass the tight selection criteria. If there are two denominators in an event, or one denominator can produce more than one type of fake lepton, both possibilities are considered separately. Such an event enters the histograms more than once, and each time scaled by the relevant fake rate, f. The equation for the fake background estimate:

$$N_{l+fake} = N_{l+Denom,i} \times \frac{f_i}{1 - f_i} \tag{5.2}$$

This equation appropriately corrects the contamination of real leptons in the denominator objects when extracted from jet trigger data.

A further correction is applied for non-triggerable fake leptons. Such a correction is necessary because denominator objects generally do not trigger the event. The full complement of lepton plus denominator object events is thus missing the case where the fake lepton triggered the event. We use the inclusive W Monte Carlo to calculate the correction by calculating a "fake yield" by applying our fake rate procedure twice. First we apply it to lepton-denominator pairs which could only trigger because of the fake lepton. The resulting yield is called $N^{MC}_{\rm nontrig}$. We then treat the sample just like data requiring the identified lepton to be triggerable giving us $N^{MC}_{\rm triggerable}$. We then scale up the fake events from data by $1 + N^{MC}_{\rm nontrig}/N^{MC}_{\rm triggerable}$, so that

$$N^{data}_{\rm total} = N^{data}_{\rm triggerable} + N^{data}_{\rm triggerable}\frac{N^{MC}_{\rm nontrig}}{N^{MC}_{\rm triggerable}}. \tag{5.3}$$

The ratios of nontriggerable to triggerable events are shown in Table 5.12.

The total background from fakes using this procedure is estimated to be 67 ± 12 events and the details are shown in table 5.13. The fake contribution of ee is factor of 7 larger than that of $\mu\mu$. There are 85% of fake coming from PHX related channel in ee.

5.5 Cross Checks

In this chapter, we present the control regions in which we have tested our data model. Each control region attempts to isolate or enhance effects that must be modeled. The control regions considered are itemized below:

- the Drell-Yan(DY) region with no $\not\!\!E_T$ cut which tests the modeling of good run / luminosity accounting, lepton ID scale factors, and non-MET related triggers.
- Same-sign which tests the contributions from photons and jets misidentified as leptons
- Low $\not\!\!E_T^{spec}$ high $\not\!\!E_T$ which test the effect of mismeasured individual leptons and jets.
- Low $\not\!\!E_T{}^{\rm sig}$ high $\not\!\!E_T^{spec}$ which tests the effect of mismeasured unclustered energy or combinations of leptons on the $\not\!\!E_T$ modeling

Table 5.12: Ratio of non-triggerable to triggerable fakes by category and region.

	Base	Base Same Sign	Base Low $\not\!\!E_T{}^{\rm sig}$
TCE TCE	0.04 ±0.00	0.04 ±0.00	0.02 ±0.00
TCE PHX	0.04 ±0.00	0.04 ±0.00	0.01 ±0.00
PHX PHX	0.03 ±0.01	0.03 ±0.01	0.02 ±0.00
e e	0.04 ±0.00	0.04 ±0.00	0.02 ±0.00
TCE CMUP	0.04 ±0.00	0.04 ±0.00	0.04 ±0.01
TCE CMX	0.04 ±0.01	0.04 ±0.01	0.00 ±0.00
TCE CMIOCES	0.22 ±0.01	0.21 ±0.01	0.25 ±0.03
TCE CMIOPES	0.18 ±0.01	0.18 ±0.01	0.19 ±0.03
PHX CMUP	0.03 ±0.00	0.03 ±0.00	0.03 ±0.01
PHX CMX	0.03 ±0.01	0.03 ±0.01	0.00 ±0.00
PHX CMIOCES	0.41 ±0.03	0.40 ±0.03	0.62 ±0.10
PHX CMIOPES	0.32 ±0.03	0.32 ±0.03	0.63 ±0.11
e mu	0.12 ±0.00	0.12 ±0.00	0.14 ±0.01
CMUP CMUP	0.05 ±0.02	0.06 ±0.02	0.09 ±0.02
CMUP CMX	0.05 ±0.02	0.06 ±0.02	0.09 ±0.02
CMUP CMIOCES	0.50 ±0.06	0.49 ±0.06	0.47 ±0.06
CMUP CMIOPES	0.54 ±0.07	0.53 ±0.07	0.59 ±0.08
CMX CMX	0.00 ±0.00	0.00 ±0.00	0.00 ±0.00
CMX CMIOCES	0.48 ±0.07	0.43 ±0.07	0.37 ±0.07
CMX CMIOPES	0.27 ±0.05	0.27 ±0.05	0.33 ±0.08
mu mu	0.29 ±0.02	0.28 ±0.02	0.30 ±0.02
TCE Trk	0.06 ±0.00	0.06 ±0.00	0.08 ±0.00
PHX Trk	0.12 ±0.01	0.11 ±0.01	0.12 ±0.01
e trk	0.08 ±0.00	0.08 ±0.00	0.10 ±0.00
CMUP Trk	0.21 ±0.02	0.21 ±0.02	0.21 ±0.02
CMX Trk	0.10 ±0.01	0.10 ±0.01	0.12 ±0.02
mu trk	0.17 ±0.01	0.17 ±0.01	0.18 ±0.01

Table 5.13: The expected yields for Wjet in the Base region.

Cut	ee	$e\mu$	$\mu\mu$	et	$\mu\ t$	$ll(\pm stat)$
TrigDil	695.7	178.1	229.2	537.3	231.5	1871.8 ± 9.1
maxΔz	687.1	150.9	222.2	524.7	226.6	1811.5 ± 8.9
m_{ll}	677.2	148.7	221.0	520.0	222.1	1789.0 ± 8.9
$\not\!\!E_T^{spec}$	48.7	47.5	5.2	18.0	9.9	129.2 ± 2.4
OS	24.5	26.3	3.5	11.7	6.6	72.7 ± 1.8
Njets<2	22.8	24.1	3.2	11.0	5.8	66.8 ± 1.7
0jet	17.6	17.9	2.5	9.2	4.5	51.7 ± 1.5

Table 5.14: Predicted and Observed Yields in the Drell-Yan Region

Category	WW	WZ	ZZ	$t\bar{t}$	DY	$W\gamma$	W+jets	Total	Data
$e\ e$	4.6	25.9	32.2	0.9	69849.0	1.2	304.1	70217.9 ± 7719.2	67962
$e\ \mu$	4.4	0.1	0.0	0.8	54.9	0.3	14.8	75.3 ± 8.7	68
$\mu\ \mu$	4.3	21.7	28.6	0.9	56363.8	0.0	195.4	56614.7 ± 5114.4	54562
e trk	4.2	8.4	10.8	0.8	21609.8	0.4	320.1	21954.6 ± 3503.5	20307
μ trk	2.6	7.5	9.7	0.5	18830.1	0.0	174.5	19024.9 ± 1821.6	18081
Total	20.2	63.5	81.3	3.9	166707.6	1.9	1008.9	167887.4 ± 17647.6	160980

We choose two separate control regions for $\not\!\!E_T$ mis-modelling because $\not\!\!E_T$ can arise due to either a single object (i.e. a lepton or a jet) being mismeasured, or due to tails in the response function of the detector as a whole. Low $\not\!\!E_T{}^{\rm sig}$ is also included in the Base region for Higgs sensitivity tests since we want to use leptonic kinematics information only for Higgs analysis. There's no event selection cuts based on $\not\!\!E_T{}^{\rm sig}$.

5.5.1 Dilepton Z-mass (DY) Control Region

Here we consider opposite-sign same-flavor dileptons in the Z mass region which is completely dominated by the Drell-Yan process. We require one trigger leg with $E_T > 20$ GeV and the other leg has $E_T > 10$ GeV. We also require $\not\!\!E_T < 20$ GeV in this region to be exclusive from Base region.

Table 5.14 and Figures 5.10 show the agreement between the expected number of events and the observed yield for each dilepton category. The systematics are not shown in the plots. The uncertainty of yields in the table include the systematics which will be described in Chapter 7. A general agreement within $\pm 1\sigma$ consistency including systematics is demonstrated. It is satisfactory for a search type analysis.

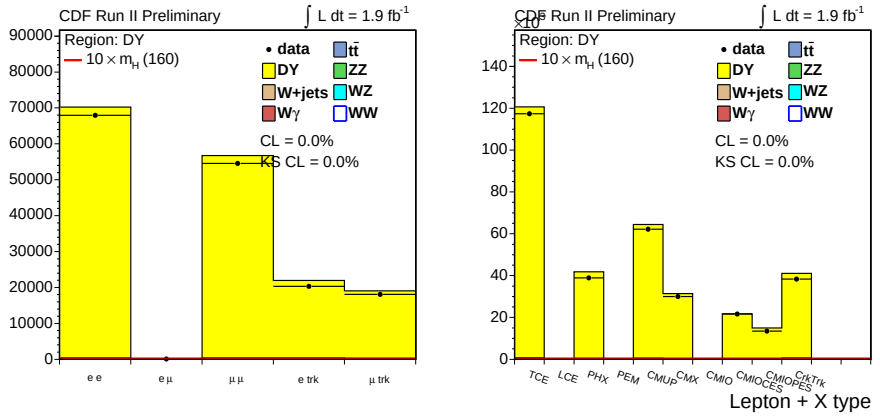


Figure 5.10: The expected and observed dilepton yields in the Drell-Yan ($Z \rightarrow l^+l^-$) control region. The right plot shows the expected and observed leptons used in the DY control region. Each event contributes two entries in the histogram.

5.5.2 Base Same Sign Control Region

As a check of the W+jets and $W\gamma$ fake backgrounds, we look at the data in the signal region with full *Base* selection, but with the opposite sign lepton requirement reversed. This region is dominated by fake leptons from jets or conversion photon. The PHX categories also has significant contributions from charge mis-id. The results are shown in Table 5.15. We find a general agreement of $\approx 1\sigma$ including the systematic. In addition there is no pattern of discrepancy in a particular channel and the kinematic distributions shown in Figures 5.11 are well modeled.

Table 5.15: Expected and observed yields in the Same-Sign region

Category	WW	WZ	ZZ	$t\bar{t}$	DY	$W\gamma$	W+jets	Total	Data
$e\ e$	2.9	1.5	0.6	0.1	2.1	26.7	22.4	56.4 ± 8.2	59
$e\ \mu$	3.5	3.1	0.3	0.2	5.5	25.1	19.8	57.5 ± 8.3	71
$\mu\ \mu$	0.0	1.1	0.2	0.0	0.1	0.0	1.7	3.0 ± 0.6	4
e trk	1.2	1.1	0.2	0.1	1.1	6.4	6.3	16.3 ± 2.0	21
μ trk	0.0	0.6	0.1	0.0	0.3	1.1	3.1	5.2 ± 0.5	6
Total	7.6	7.3	1.3	0.4	9.1	59.3	53.4	138.4 ± 19.2	161

5.5.3 Low $\not\!\!E_T^{spec}$ High $\not\!\!E_T$ Control Region

As a cross check of the $\not\!\!E_T$ modeling, we look in the region of $\not\!\!E_T^{spec}$ between 15 GeV (10 GeV $e\mu$) and the *Base* region requirement of 25 GeV (15 GeV $e\mu$) but with $\not\!\!E_T$

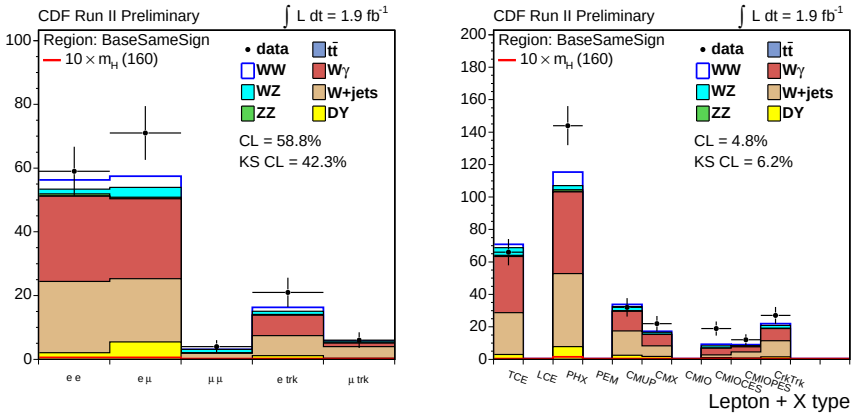


Figure 5.11: The expected and observed dilepton and lepton yields in the Base Same Sign control region.

greater than 25 (15 GeV $e\mu$). This region is dominated by Drell-Yan with fake high $\not\!\!E_T$ due to mismeasured single leptons or jets. $\not\!\!E_T$ will point to mis-measured leptons or jets direction. In this region , we find good agreement between the data and our model as shown in Table 5.16 and Figures 5.12. There is 20% deviation of observed and expected yields in this region. We assign the $\not\!\!E_T$ uncertainty to be 20% for Drell-Yan background.

Table 5.16: Expected and observed yields in the Base Low $\not\!\!E_T^{spec}$ High $\not\!\!E_T$ region

Category	WW	WZ	ZZ	$t\bar{t}$	DY	$W\gamma$	W+jets	Total	Data
e e	3.7	1.0	1.0	0.4	33.7	2.4	4.6	46.9 ± 7.4	40
e μ	6.0	0.3	0.1	0.5	31.0	1.7	3.4	42.9 ± 7.0	31
μ μ	2.8	0.7	0.8	0.4	28.5	0.0	0.9	34.1 ± 6.3	33
e trk	3.1	0.5	0.4	0.3	35.4	0.5	2.9	43.2 ± 7.8	38
μ trk	1.8	0.3	0.3	0.2	7.8	0.1	1.0	11.5 ± 1.8	9
Total	17.4	2.6	2.7	1.8	136.5	4.7	12.8	178.6 ± 30.2	151

5.5.4 Low $\not\!\!E_T{}^{\rm sig}$ high $\not\!\!E_T^{spec}$ Control Region

As a second cross check of the $\not\!\!E_T$ simulation and in order to assess a systematic on the Drell-Yan background, we apply a $\not\!\!E_T{}^{\rm sig} \equiv \not\!\!E_T/\sqrt{\sum E_T} < 2.5$ GeV$^{\frac{1}{2}}$ cut to the *Base* region. The resulting region is populated primarily by Z+jets where Drell-Yan gains large boost in the transverse plane to balance the real emission jets. It will deposit large EM and Hadronic energy in the calorimeter and gives larger $\not\!\!E_T$ uncertainty. The $\not\!\!E_T$ uncertainty is proportional to $\sum E_T$ which is the transverse energy scalar sum for all

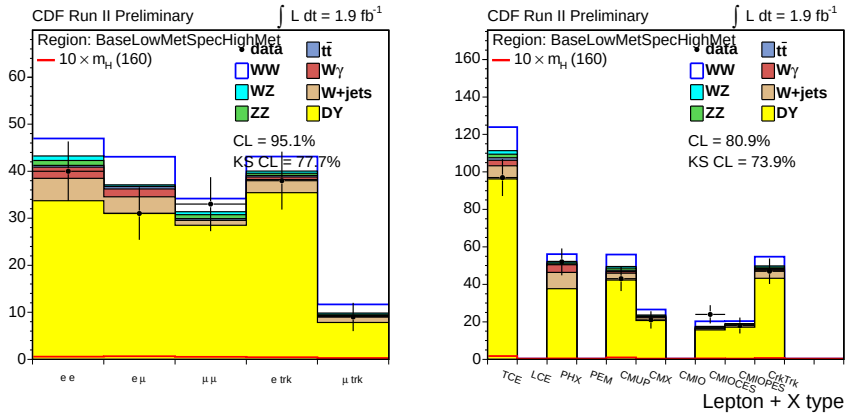


Figure 5.12: Expected and observed dilepton and lepton yields in the Base Low $\not\!\!E_T^{spec}$ High $\not\!\!E_T$ control region.

calorimeter towers. Drell-Yan events with $\not\!\!E_T$ due to lost energy from a single lepton or jet is removed by the $\not\!\!E_T^{spec}$ cut. In the end, this region is to test the mis-measurement of the unclustered energy or a combination of leptons and jets. We also find good agreement between the data and our model as shown in Table 5.17 and Figures 5.13.

Table 5.17: Expected and observed yields in the Low $\not\!\!E_T{}^{\rm sig}$ region

Category	WW	WZ	ZZ	$t\bar{t}$	DY	$W\gamma$	W+jets	Total	Data
$e\ e$	2.9	0.5	0.5	0.2	17.8	2.4	3.4	27.7 ± 4.0	26
$e\ \mu$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0 ± 0.0	0
$\mu\ \mu$	2.3	0.4	0.5	0.2	10.6	0.0	0.5	14.4 ± 2.4	13
e trk	2.5	0.2	0.2	0.2	6.5	0.6	1.0	11.3 ± 1.5	11
μ trk	1.4	0.1	0.2	0.1	3.8	0.1	0.6	6.3 ± 0.9	5
Total	9.1	1.3	1.4	0.8	38.7	3.1	5.4	59.7 ± 8.6	55

5.6 Candidate Events in Base Region

The resulting background and signal expectations for the Base region are shown in Table 5.18 for the backgrounds and 5.5 for the various Higgs masses. The expected WW yield is 251 events. The expected distributions for these events are shown in Figures 5.14. For comparison, a 160 GeV/c^2 Higgs with yields scaled up by a factor 10 is also shown. we stack the backgrounds on top of each other, and then overlay the plot depicting 10× the Higgs yield for a Higgs mass of 160 GeV/c^2. The observed and

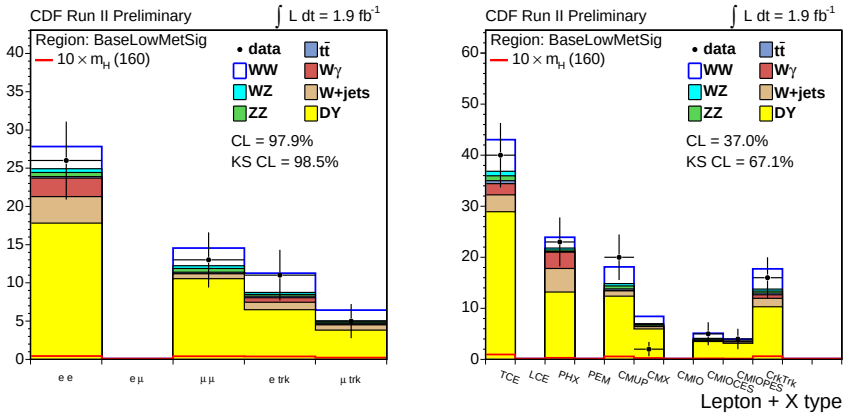


Figure 5.13: Expected and observed dileptona and lepton yields in the Base Low$\not\!\!E_T{}^{\rm sig}$ control region.

expected yields in Base region are shown in Table 5.18.

Table 5.18: The Base region yields in the five dilepton categories.

Category	WW	WZ	ZZ	$t\bar{t}$	DY	$W\gamma$	W+jets	Total	Data
e e	46.6	5.3	5.6	2.9	26.6	27.2	22.8	139.5 ± 11.8	144
e μ	110.1	3.2	0.4	7.0	22.5	23.8	24.1	191.1 ± 16.6	191
μ μ	36.0	4.1	4.8	2.7	17.6	0.0	3.1	70.1 ± 6.1	58
e trk	37.8	2.6	2.4	2.6	10.3	6.5	10.9	73.9 ± 5.6	80
μ trk	20.6	1.6	1.7	1.5	5.3	1.1	5.8	38.3 ± 3.0	49
Total	251.0	16.9	15.0	16.8	82.2	58.5	66.6	507.0 ± 41.1	522

Category	WW	WZ	ZZ	tt	DY	$W\gamma$	W+jets	Total	Data
TCE TCE	25.1	2.6	3.4	1.9	12.6	5.1	3.3	54.1 ± 4.5	53
TCE PHX	17.9	2.3	2.3	0.9	10.9	14.7	12.7	61.5 ± 5.5	70
PHX PHX	3.5	0.4	0.4	0.1	3.1	7.4	6.8	21.8 ± 2.4	21
TCE CMUP	32.6	0.8	0.2	2.5	7.5	3.3	3.7	50.6 ± 4.5	45
TCE CMX	13.7	0.4	0.1	1.0	2.9	1.4	1.6	20.9 ± 1.8	20
TCE CMIOCES	23.9	0.6	0.1	1.7	3.7	1.3	1.3	32.6 ± 3.0	30
TCE CMIOPES	11.9	0.3	0.0	0.6	1.7	0.7	2.0	17.3 ± 1.7	11
PHX CMUP	11.1	0.4	0.1	0.6	2.3	7.3	8.9	30.6 ± 2.9	43
PHX CMX	5.0	0.2	0.0	0.2	1.6	4.5	4.4	16.0 ± 1.6	19
PHX CMIOCES	7.5	0.3	0.0	0.3	1.5	3.0	1.0	13.7 ± 1.5	12
PHX CMIOPES	4.2	0.1	0.0	0.1	1.3	2.3	1.3	9.4 ± 1.1	11
CMUP CMUP	6.9	0.7	1.0	0.6	3.2	0.0	0.4	12.9 ± 1.1	17
CMUP CMX	6.3	0.7	0.8	0.5	2.5	0.0	0.4	11.2 ± 1.0	6
CMUP CMIOCES	10.0	1.1	1.4	0.8	3.8	0.0	0.6	17.8 ± 1.6	20
CMUP CMIOPES	5.1	0.6	0.6	0.3	3.7	0.0	0.9	11.2 ± 1.2	5
CMX CMX	1.3	0.1	0.2	0.1	0.7	0.0	0.1	2.6 ± 0.2	0
CMX CMIOCES	4.3	0.5	0.7	0.3	1.8	0.0	0.3	7.9 ± 0.7	6
CMX CMIOPES	2.1	0.2	0.3	0.1	1.9	0.0	0.3	5.0 ± 0.5	4
TCE Trk	28.6	1.9	1.9	2.2	7.6	3.0	6.1	51.3 ± 3.9	48
PHX Trk	9.2	0.7	0.5	0.4	2.6	3.5	4.8	21.7 ± 1.8	32
CMUP Trk	14.2	1.1	1.2	1.1	3.4	0.8	3.8	25.7 ± 2.1	40
CMX Trk	6.4	0.5	0.5	0.4	1.9	0.4	2.0	12.1 ± 1.0	9
Total	251.0	16.9	15.7	16.8	82.2	58.5	66.6	507.7 ± 41.1	522

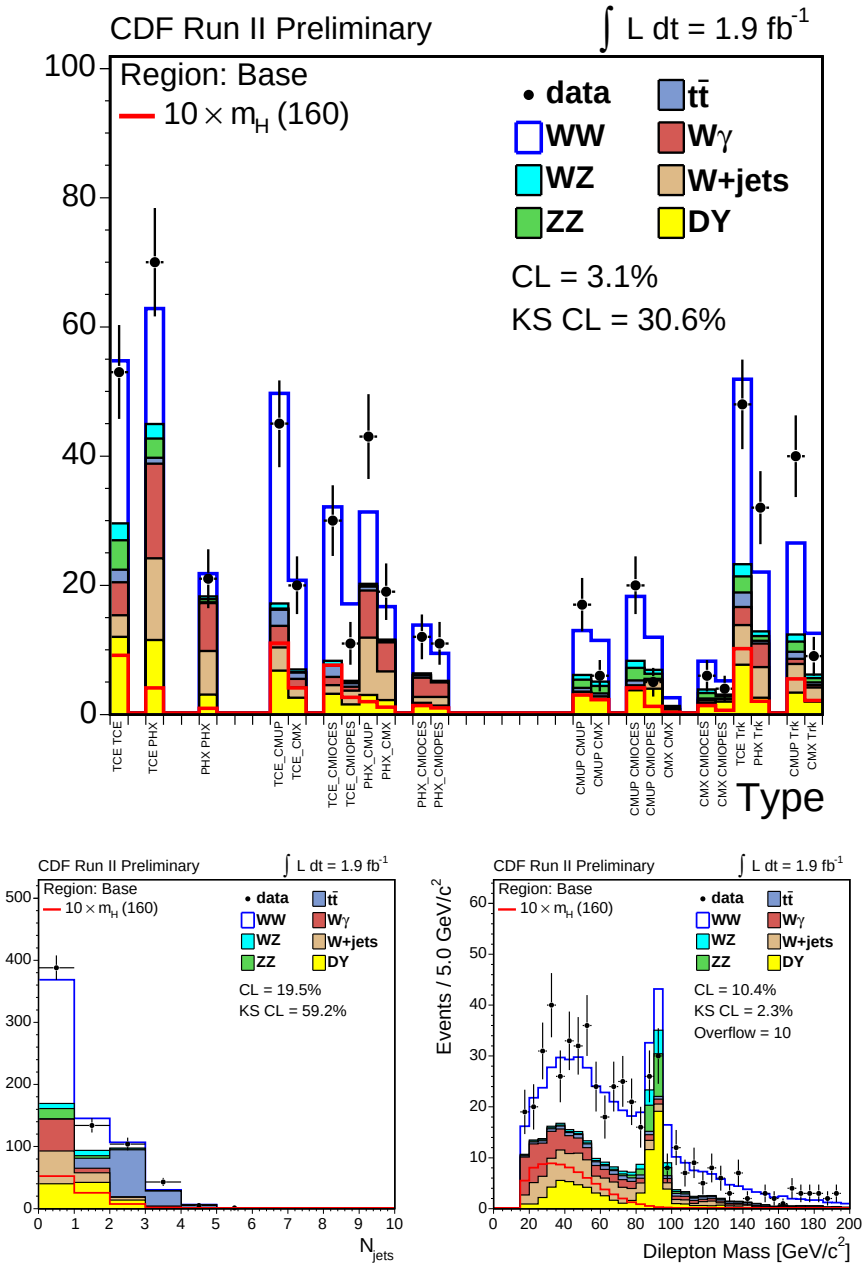


Figure 5.14: The kinematic distributions of signal and backgrounds in the base region.

Chapter 6

Matrix Element Method

The Matrix Element Method is a generic way to calculate event probabilities for each candidate event by using all available kinematic information. These probabilities can be used as a discriminant in this analysis, but can also be used directly in maximum likelihood fits. The probabilities are calculated using the leading order theoretical predictions for the differential cross-section for each contribution. As all of the information is available in the fit, and the likelihood is calculated analytically, we can easily implement a range of models covering a large parameter space without having to re-optimize the selection of the fit variables.

The Matrix Element Method has been applied to top mass measurements [67] and the single top search [68]. The technical difficulty here is the missing information of the neutrino final states. The differential cross-section of given measured kinematics are calculated through integrating over missing information. Monte Carlo integration techniques are used to enable the numerical calculation. To reduce the computing time, importance sampling is used as base integration and cross checked with the Vegas integration [69]. The Matrix Element function used in this analysis is from MCFM v3.4.5[70]. MCFM provides both LO, and NLO cross-section calculations for all relevant diboson and Higgs processes in pp or $p\bar{p}$ colliders. A potential improvement for this method could be doing NLO calculation, but this is not addressed in this thesis.

The structure of this chapter is described as below. The equations used to calculate event probabilities are explained. It follows the construction of transfer function: resolution, efficiency and the p_T spectrum of hard interaction. The likelihood ratio discriminators are constructed as the final observables. The cross check of event probability calculation is demonstrated by comparing Monte Carlo to real data. Finally, a

comparison of the sensitivity improvement compared to conventional cut based analysis is presented.

6.1 Event Probability Calculation

The Matrix Element technique directly calculates the event probability, $P(x_{obs};\alpha)$, where α is a set of physics parameters of the specific model and x_{obs} are the measured quantities. In the case of Standard Model Higgs search, α is (m_H, Γ_H), where m_H is Higgs mass and Γ_H is Higgs width. There are eight observables x_{obs} representing all the leptonic kinematic information: lepton momentum $\vec{l}^{+}$, $\vec{l}^{-}$ and missing transverse momentum, $\not\!\!E_{Tx}$ and $\not\!\!E_{Ty}$.

It should be noted that additional information such as the number of jets produced and the total visible energy ($\sum E_T$) may further differentiate $q\overline{q} \to WW$ from $gg \to H \to WW$, but they can suffer from significant QCD uncertainties [71]. In fact $\sum E_T$ is poorly modeled in CDF and studied in the section 7.3.2. For this reason we deliberately do not use hadronic information at all but just use leptonic information and $\not\!\!E_T$.

The event probability density is

$$P(x_{obs};\alpha) = \frac{1}{<\sigma(\alpha)>}\int \frac{d\sigma_0(y;\alpha)}{dy}\epsilon(y)G(x_{obs},y)dy, \tag{6.1}$$

where y are the true values of the observables, $\frac{d\sigma_{LO}}{dy}$ is the parton-level differential cross-section differential in those observables, $\epsilon(y)$ is detector acceptance and efficiency function and $G(x_{obs},y)$ is the transfer function representing the detector resolution. Equation (6.1) integrates over all possible true values of the observables, y, consistent with the measured quantities x_{obs}. The constant $<\sigma(\alpha)>$ normalizes the total event probability to unity, i.e.

$$\int_{x_{obs}\in V_{acceptance}} P(x_{obs};\alpha)dx_{obs} = 1. \tag{6.2}$$

Due to the two neutrinos in the final state, we have missing information that can not be measured. The differential cross-section, $\frac{d\sigma_0}{dy}$, must be calculated by integrating over the missing information (the energy and directions of the neutrinos under the constraint of the measured $\not\!\!E_{Tx}$ and $\not\!\!E_{Ty}$). As the theoretical matrix element contains the information about the W or Z resonances, the corresponding mass constraints are implicitly used. For example in $gg \to H \to WW^* \to l^+l^-\nu\bar{\nu}$, the parton-level differential cross-section as a function of the observables y is

$$\begin{aligned} \frac{d\sigma_0(y;\alpha)}{dy} &= \int \frac{\sum_{ij} f_i(x_1,Q^2) f_j(x_2,Q^2) |M_{ij}(\alpha)|^2}{4\sqrt{(q_1\cdot q_2)^2 - m_{q_1}^2 m_{q_2}^2}} \\ &\times \quad dx_1 dx_2 d\Phi_N(z)\delta(y-z)k_T(k_x,k_y)dk_x dk_y, \end{aligned} \tag{6.3}$$

where $f_{i(j)}(x_{1(2)})$ is the parton density function, $x_1(2)$ is the parton energy fraction and Q^2 is the momentum transfer. $d\Phi_N(x)$ is N body decay phase space. $k_T(k_x,k_y)$ is the transverse momentum distribution accounting for the proton intrinsic k_T distribution. It doesn't change the rate of total cross-section by construction $\int k_T(k_x,k_y)dk_x dk_y = 1$. M_{ij} is the matrix element which determines the dynamics of the physics model. The delta function, $\delta(y-z)$, constrains the kinematics to the input observables y when integrating over the full phase space $d\Phi_N$. For example, the four body decay phase space is:

$$d\Phi_4 = \delta^4(p_1 + p_2 - \sum l_i) \prod_i \frac{d^3 l_i}{(2\pi)^3 2E_{l_i}}. \tag{6.4}$$

Along with the two p_T dimensions and the two parton energy fractions x_1, x_2, there are a total of 16 integration degrees of freedom. However, there are 8 constraints from $\delta(y-x)$ and another four constraints from energy-momentum conservation relation in $d\Phi_4$. The final number of degrees of freedom for $d\sigma/dy$ calculation is 16-12=4. The details of calculations are shown in the Appendix D.

Resolution Function

The resolution function is a Gaussian smearing applied to the true lepton energy to simulate measured lepton energy. The electron and muon p_T are smeared by the CDF detector Parameters[72].

$$\frac{\sigma(E)}{E} = \frac{13.5\%}{\sqrt{E_T}}(TCE) \tag{6.5}$$

$$\frac{\sigma(E)}{E} = \frac{14.4\%}{\sqrt{E}} \oplus 0.7\%(PHX) \tag{6.6}$$

$$\frac{\sigma(p_T)}{p_T} = \frac{0.15\%}{p_T}(Muon, CrkTrk) \tag{6.7}$$

The $\not\!\!E_T$ resolution is a function of $\sum E_T$ and follows the relationship $\sigma(\not\!\!E_T) = 60\%\sqrt{\sum E_T}$. To avoid using $\sum E_T$ information, a fixed Gaussian smearing with width 4.6 GeV is applied to $\not\!\!E_T$ in the x and y direction independently. This function was determined from

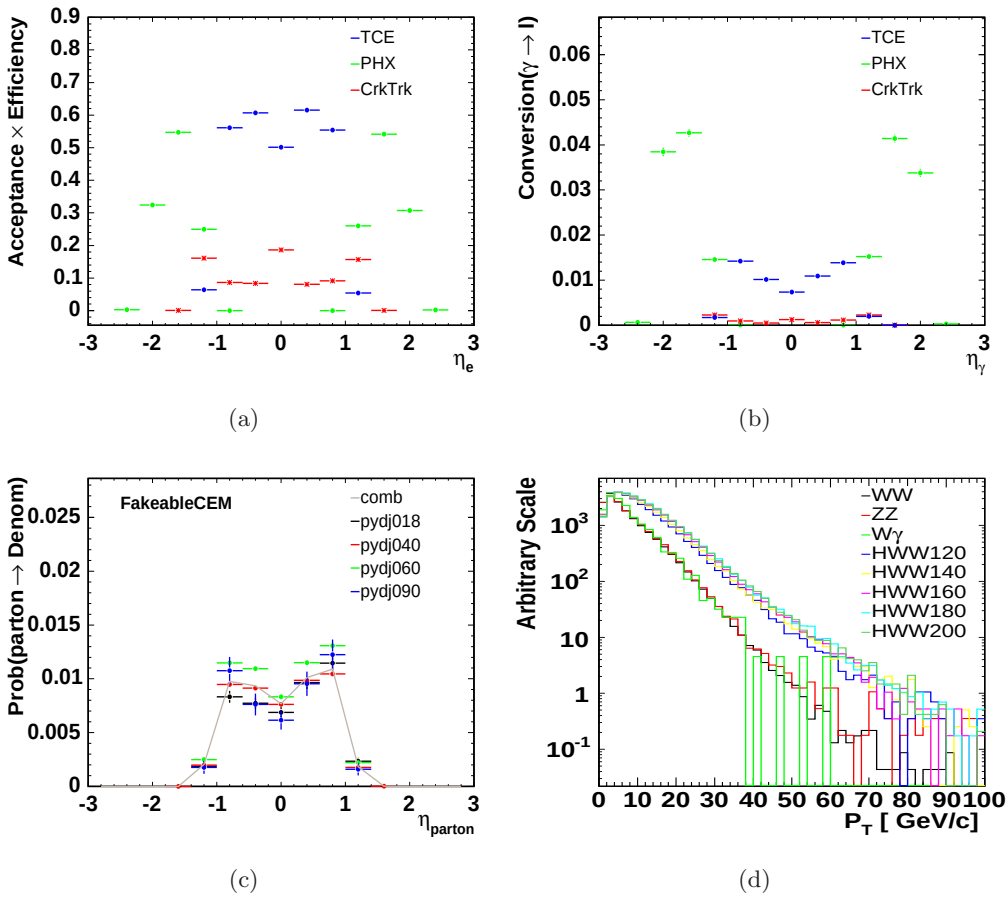


Figure 6.1: (a) Lepton efficiency function as a function of pseudo-rapidity η. (b) Photon conversion efficiency. (c) Probability for a jet to be faked as a denominator. (d) k_T of processes for interests extracted from PYTHIA.

the core of the Drell-Yan $\not\!\!E_T$ distribution. To simplify the calculation, the modeling of a photon or a jet smeared to be a reconstructed lepton is treated the same as the real electron or muon does.

Efficiency Function

The efficiency function is the probability for a parton-level object with momentum p being reconstructed to a lepton with momentum q. The resolution function takes into account the smearing of transverse energy but assumes the angule is measured perfectly. The parton level lepton p equals q. Figures 6.1(a) shows the lepton efficiency parameterized with pseudo-rapidity η extracted from inclusive W Monte Carlo. The efficiency function in Equation 6.1 can be pulled out from the integral when calculating

WW event probability,

$$P_{WW}(x_{obs};\alpha) = \frac{\epsilon(\eta_{1,obs})\epsilon(\eta_{2,obs})}{<\sigma_{WW}>}\int\frac{d\sigma_{WW}(y;\alpha)}{dy}G(x_{obs},y)dy. \quad (6.8)$$

Figures 6.1(b) shows the probability for a photon being constructed as an electron also paramatrized as η. The formula to calculate $W\gamma$ event probability can be written as:

$$P_{W\gamma}(x_{obs};\alpha) = \frac{\epsilon(\eta_{1,obs})\epsilon_{\gamma\to l}(\eta_{2obs})}{<\sigma_{W\gamma}>}\int\frac{d\sigma_{W\gamma}(y;\alpha)}{dy}G(x_{obs},y)dy + 1 \leftrightarrow 2. \quad (6.9)$$

The permutation of two leptons are considered since both have chances to be photon conversions. Figure D.1 shows the probability for a parton-level jet being reconstructed to a denominator object as a function of pseudo-rapidity η. The fake rate measured from data, $\epsilon_{D\to l}(p_{T,obs})$, is multiplied to get the final efficiency for a jet reconstructed as lepton. The Wjet event probability is written as

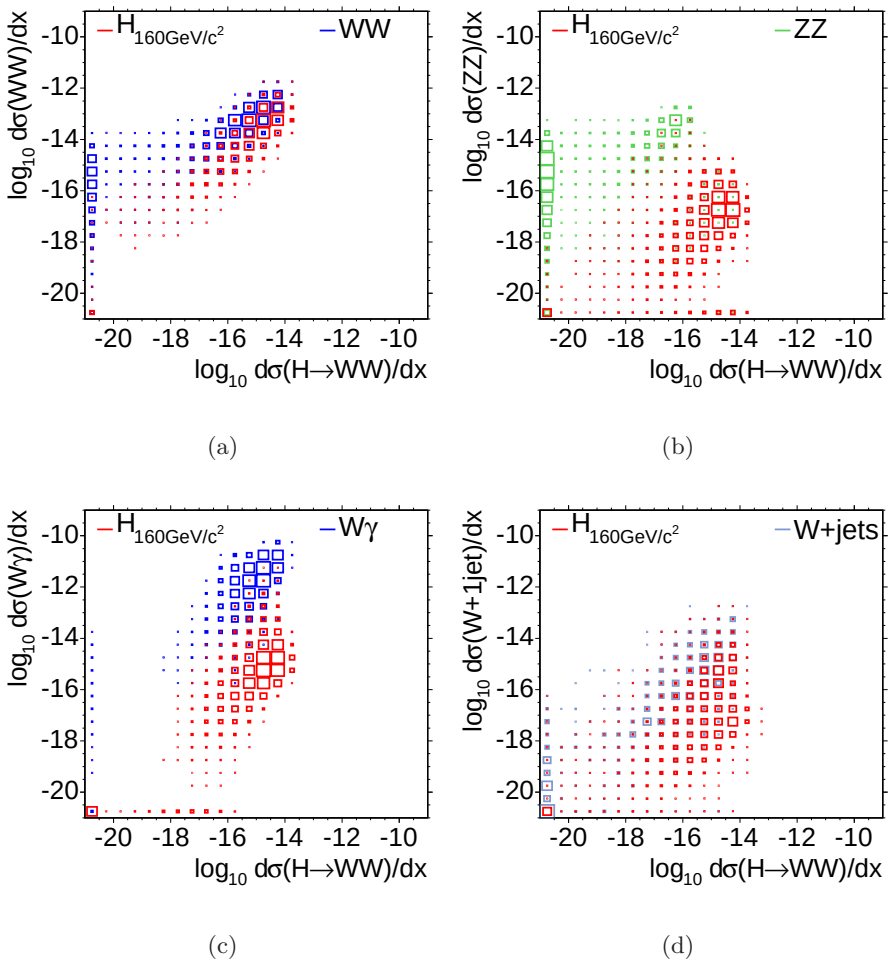
$$P_{Wjet}(x_{obs};\alpha) = \frac{\epsilon(\eta_{1obs})\epsilon_{j\to D}(\eta_{2obs})\epsilon_{D\to l}(p_{T2,obs})}{<\sigma_{W\mathrm{jet}}>}\int\frac{d\sigma_{WW}(y;\alpha)}{dy}G(x_{obs},y)dy + 1 \leftrightarrow 2. \quad (6.10)$$

k_T Function

The k_T model is extracted from Pythia Monte Carlo for each process to account for the transverse recoil effect. It therefore accounts for the intrinsic k_T of the proton and the Pythia parton shower model prediction of soft QCD effects. Figure 6.1(d) shows that gluon-gluon fusion production is harder than qq' production.

Performance of Event Probabilities

There are total five event probabilities calculated. Each event can be characterized with five variables. To appreciate the discrimination power, the correlation of differential cross-section of each background v.s. Higgs are studied. Figure 6.2(a) shows WW v.s. Higgs with clear distinction. There are phase spaces in WW sample distinguishable and indistinguishable from Higgs. The ZZ v.s. Higgs has better discrimination as shown in Figure 6.2(b). Without $d\sigma(W\gamma)$, $W\gamma$ and Higgs are hard to discriminate due to the fact that final state radiation photon also has small opening angle and low dilepton mass. Figure 6.2(c) shows $d\sigma(W\gamma)$ has successfully separate two samples. The Wjet to Higgs discrimination is less powerful. It is caused by the simplified procedure of $d\sigma(W\mathrm{jet})$ calculation. The event probability can not describe background model well and the consequence is to weaken the discrimination power.



(a) (b)

(c) (d)

Figure 6.2: Parton level differential cross-section scatter plots for Higgs at 160 GeV/c vs other processes.

6.2 Likelihood Ratio Discriminator

In principle, we could, and may still at some future time, use the Matrix Element in a Maximum Likelihood fit for either the fractional or absolute yields in the data. In practice, this would require us to have complete models of all the backgrounds, and fully understand the pull distributions of the Maximum Likelihood fit. At this time we do not have a complete model of the data; we do not calculate probabilities for Drell-Yan or WZ for technical reasons, and have only LO Matrix Element calculations implemented. We thus decided to use a simpler approach by following the strategy used in the single top analysis [68] where they use the event probability calculated with a Matrix Element technique to construct a likelihood ratio (LR) discriminator. Using this

discriminator in a 1-dimensional template fit, we do not need a complete model of the data. It is sufficient to model only the largest components and any mis-modeling in the analytic calculation will only result in a less than optimal sensitivity. Furthermore, the discriminant technique allows us to use MC@NLO to make a template that is valid to NLO for the WW component, where the shape is most critical. The discriminator is defined as :

$$LR = \frac{P_s}{P_s + \sum_i k_{bi} P_{bi}}, \tag{6.11}$$

where k_{bi} are the relative ratio of expected contributions of each background and satisfy $\sum k_{bi} = 1$. The calculation of P_s is a function of Higgs mass so that likelihood ratio shape depends on m_H. This is true for both signal and background templates of LR.

The Higgs LR shows a clear discrimination from other backgrounds. Figure 6.3 shows the kinematic distributions of $m_H = 160$ GeV/c^2 after requiring LR > 0.9.

6.3 Two Dilepton Categories: high S/B and low S/B

As the backgrounds from Drell-Yan, $W\gamma$ and W+jets differ significantly between the various reconstructed dilepton modes, we divide the data into two categories of high and low S/B:

high S/B	low S/B
TCE TCE	TCE PHX
TCE CMUP	PHX PHX
TCE CMX	PHX CMUP
TCE CMIOCES	PHX CMX
TCE CMIOPES	PHX CMIOCES
CMUP CMUP	PHX CMIOPES
CMUP CMX	CMUP CMIOPES
CMUP CMIOCES	CMX CMIOPES
CMX CMX	PHX CrkTrk
CMX CMIOCES	
TCE CrkTrk	
CMUP CrkTrk	
CMX CrkTrk	

These categories as based on the $N_s/\sqrt{N_s + N_b}$ curve as a function of LR cut as shown in Figure 6.4. The yields for the high and low S/B categories are shown in table 6.1 The resulting LR distributions for Higgs masses 160 GeV/c^2 are shown in Figure 9.5.

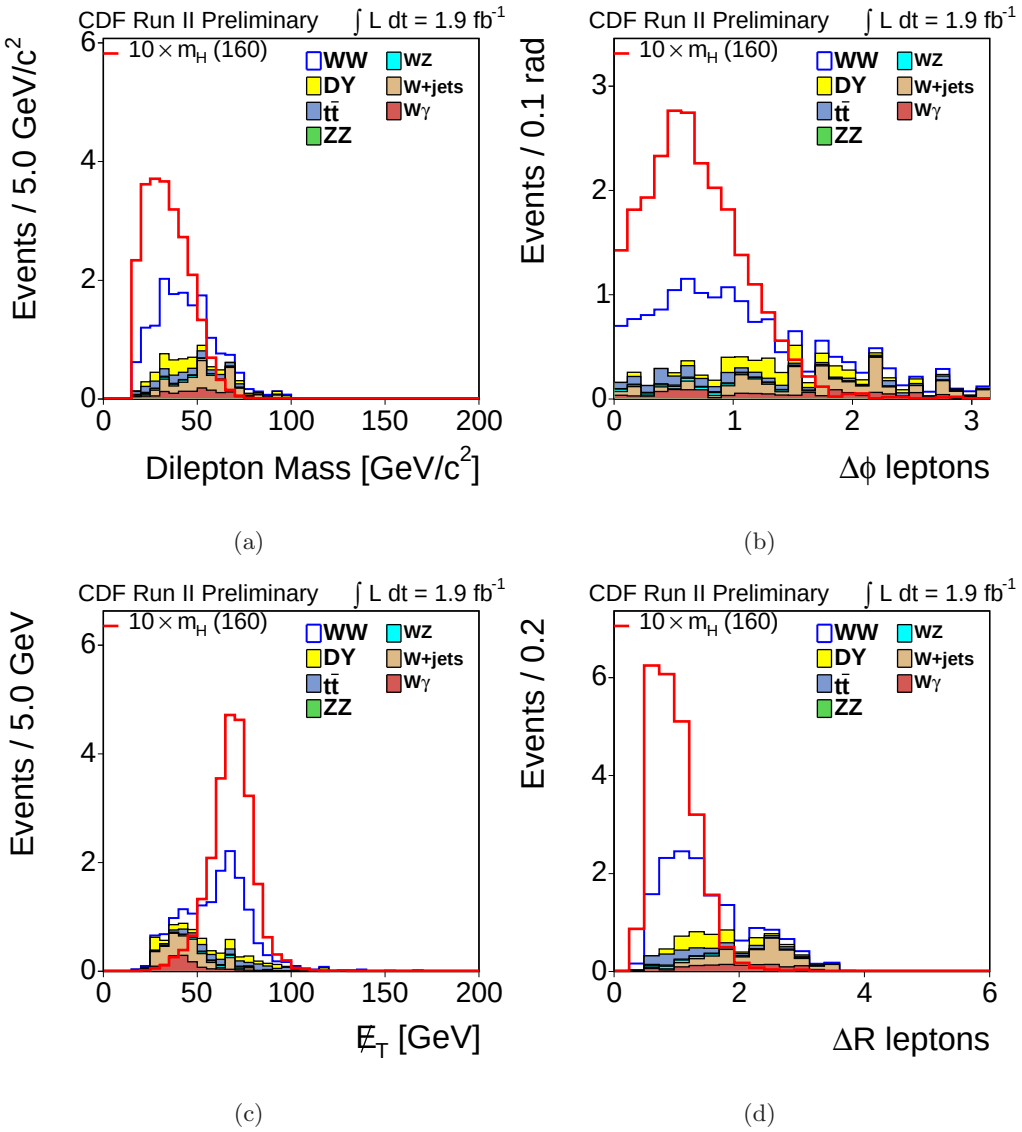
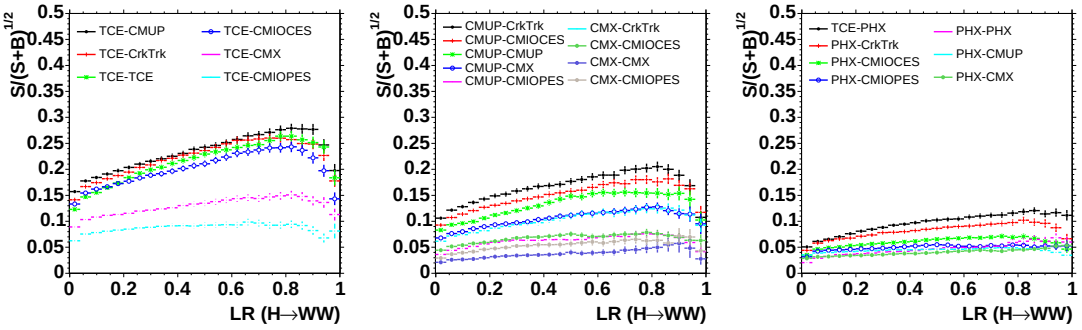


Figure 6.3: Kinematic distributions which pass LR($m_H = 160$ GeV/c^2) > 0.9 cut.

6.4 Cut Based Analysis vs Matrix Element Methods

The performance using a cut based analysis is presented in this section. A set of cuts are applied and the significance is tested with remaining one or two variables. The sensitivity is defined as 95% C.L. limit production cross-section evaluated through Bayesian statistics. The systematics is described in detail in chapter 7.



(a) (b) (c)

Figure 6.4: The $S/\sqrt{S+B}$ as a function of Likelihood Ratio Discriminator cut for Higgs mass 160 GeV/c^2 with luminosity $1.9fb^{-1}$. (a) TCE related channels (b) Muon related channels (c) PHX related channels.

Table 6.1: Expected yields divided in the low and high S/B categories in the Base region.

Category	WW	WZ	ZZ	$t\bar{t}$	DY	$W\gamma$	W+jets	Total($\pm$ stat.)
HighSoverB	185.4	11.6	10.9	13.8	53.3	15.9	25.6	320.7 $\pm$ 2.2
LowSoverB	65.7	5.3	4.1	3.1	28.9	42.6	41.1	192.3 $\pm$ 2.1
Sum	251.1	16.9	15.0	16.8	82.2	58.5	66.6	507.0 $\pm$ 3.0

	Higgs Mass (GeV/c^2)									
Category	110	120	130	140	150	160	170	180	190	200
HighSoverB	0.3	1.1	2.4	3.9	5.1	6.3	6.2	5.0	3.6	2.8
LowSoverB	0.1	0.3	0.6	0.9	1.3	1.4	1.4	1.2	0.9	0.7

6.4.1 Bayesian's Limit

We extract the Higgs production cross section by doing histogram fit of signal distribution over background distributions. As input distributions we provide 2x9 histograms of the Higgs likelihood ratio LR_{Higgs}, two for low and high S/B, and 9 for the different Higgs hypotheses, WW, $W\gamma$, W+jets, WZ, ZZ, ttbar, DY and low dilepton mass DY. A joint likelihood of oberserved events is calculated based on Poisson statistics through each bin. A 95% Confidence Level(C.L.) limit is set through Bayesian approach since we have not seen a significant excess in the data [73].

Given N channels and n_k observed events in the kth channel, k=1,2,...,N, the

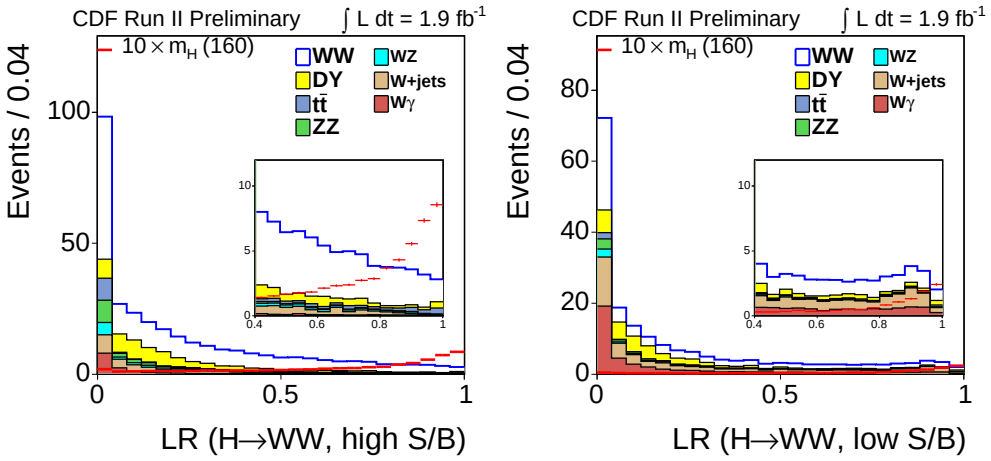


Figure 6.5: Likelihood Ratio Discriminator for Higgs mass 160 GeV/c^2.

Poisson probability of obtaining the observed result is

$$\prod_{k=1}^{N} \frac{e^{-(s\epsilon_k+b_k)}(s\epsilon_k+b_k)^{n_k}}{n_k!}, \tag{6.12}$$

where s represents the parameter of interest (the cross-section) and ϵ_k and b_k are the acceptance and expected background for the kth channel, respectively. All ϵ_k and b_k have uncertainties and are considered "nuisance parameter". In the Bayesian approach, they are assigned priors, $\pi(\epsilon_1, b_1, ..., \epsilon_N, b_N)$, which may be correlated. The marginalized posterior, $p(s)$, for s becomes

$$p(s) = \frac{\pi(s)}{N_0} \int_{(2N)} \pi(\epsilon_1, b_1, ..., \epsilon_N, b_N)[\prod_{k=1}^{N} \frac{e^{-(s\epsilon_k+b_k)}(s\epsilon_k+b_k)^{n_k}}{n_k!}]d\epsilon_1 db_1 ... d\epsilon_N db_N, \tag{6.13}$$

where $2N$ marginalization integrals are performed, and $\pi(s)$, the prior for s, is assumed for now to be independent of the joint nuisance prior. The normalization constant N_0 is given by

$$1 = \int_0^{\inf} p(s)ds \tag{6.14}$$

A flat prior is chosen for s: $\pi(s) = 1$ for $s \geq 0$ and $\pi(s) = 0$ for $s < 0$. The Bayesian 95% C.L. limit is the cross-section, $s_{95\%}$, which satisfies the equation

$$1 - \int_0^{s_{95\%}} p(s)ds = 0.05. \tag{6.15}$$

6.4.2 Sensitivity Improvements

To assess the sensitivities, we generate many MC experiments to calculate the expected limits. We compare the LR variable sensitivities to other observables. First, a previous analysis [4] at CDF is re-implemented. Second, two dimensional discriminators are studied at Base selection. All discussions are shown explicitly with Higgs mass 160 GeV/c^2 sample as example. The final sensitivity comparison is tested on Higgs mass range $110 < m_H < 200$ GeV/c^2.

Figure 6.6 shows the azimuthal angle between two leptons passing event selection cuts described in the first Higgs search results via WW^* channel at CDF II [4]. A set of mass dependent cuts is chosen which optimize $S/\sqrt{S+B}$ simultaneously in multiple observables except azimuthal angle of two leptons $\Delta\Phi_{l^+l^-}$, It is used to extract the signal yields. The summary of event selection cut is listed below:

- $p_{T1} + p_{T2} + \not\!\!E_T < m_H$
- $m_{ll} < m_H/2 - 5$ GeV/c^2
- $\not\!\!E_T > m_H/4$ GeV
- $\not\!\!E_T > 50$ GeV or $\Delta\phi(\vec{\not\!\!E}_T, \text{lepton}, \text{jet}) > 0.35$

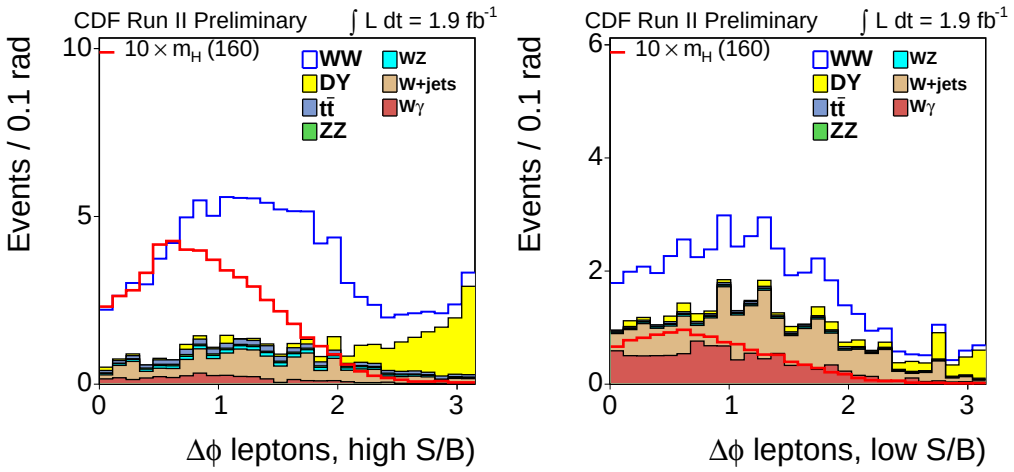


Figure 6.6: The azimuthal angle between two leptons after tight event selection cuts.

Instead of making tight cuts on each observable, we can also use two dimensional discriminators to extract Higgs yields. Figure 6.7 shows the two dimensional scatter plots

between $\Delta\Phi(l^+, l^-)$ and $M(l^+, l^-)$ in the Base region. The collinear behavior of dilepton decaying from Higgs prefer a low $\Delta\Phi(l^+, l^-)$ and m_{ll} distribution. A slice cut on dilepton mass throw information away such that we can expect improvement of sensitivity by performing a two dimensional fit.

To see the composition of backgrounds in each bin, the two dimensional histogram is re-binned to 10×5 along $m_{ll} \times \Delta\Phi_{ll}$ dimension. Each bin is projected onto one dimensional histogram as shown in Figure 6.7(c)(d). Every 5 bin is a $\Delta\Phi(ll)$ distribution between 20 GeV/c^2 segmentation of m_{ll}. The first 20 bins relate to $m_{ll} < 60$ GeV/c^2 and show a distinction of Higgs from backgrounds in $\Delta\Phi$. Bin 20-25 relate to $60 < m_{ll} < 80$ GeV/c^2 and show less discrimination along $\Delta\Phi$. The one dimensional variable analysis projects this information into $0 < m_{ll} < 75$ GeV/c^2 bin.

The other two dimensional discriminator is shown in Figure 6.8. It exploits $\Delta R(l^+, l^-) = \sqrt{\Delta\Phi^2 + \Delta\eta^2}$ and $\not\!\!E_T^{spec}$ in the Base region. $\Delta R(l^+, l^-)$ contains the information other than azimuthal angle as the η direction is included as well. $\not\!\!E_T^{spec}$ is used instead of m_{ll} based on the idea that $\not\!\!E_T$ distribution of Higgs has an upper bound due to Higgs mass. The irreducible background WW and ZZ tend to have a larger $\not\!\!E_T$ tail. It is interesting to see the performance compared to $\Delta\Phi$ vs m_{ll}.

6.4.3 Summary

Figure 6.9(a) shows the 95% C.L. production limit with respect to Standard Model production cross section of Higgs. LR variables give the best sensitivity. The expected observed limit at $m_H = 160$ GeV/c^2 is 3. To compare the different methods, the effective luminosity ratio is calculated with respect to LR variable as shown in Figure 6.9(b). The effective luminosity ratio equals to the square of cross section limit $(\sigma_i/\sigma_{LR})^2$.

In the case of $m_H = 160$ GeV/c^2, the effective accumlated luminosity gains are 33% and 40% due to adding new leptons and using LR discriminator, respectively.

The effective accumlated luminosity gains due to adding new leptons are generally 30%$\sim$40%. The effective luminosity gains by doing $\phi_{ll} - m_{ll}$ fit instead of ϕ_{ll} fit are 0%‘ $\sim$20% while it is 13% to 50% by doing $R_{ll} - \not\!\!E_T^{spec}$ fit. The effective luminosity gains by using likelihood ratio discriminator gves the best improvement which is 70% $\sim$ 150%. The total gains compared to the old method are 100% $\sim$ 240%.

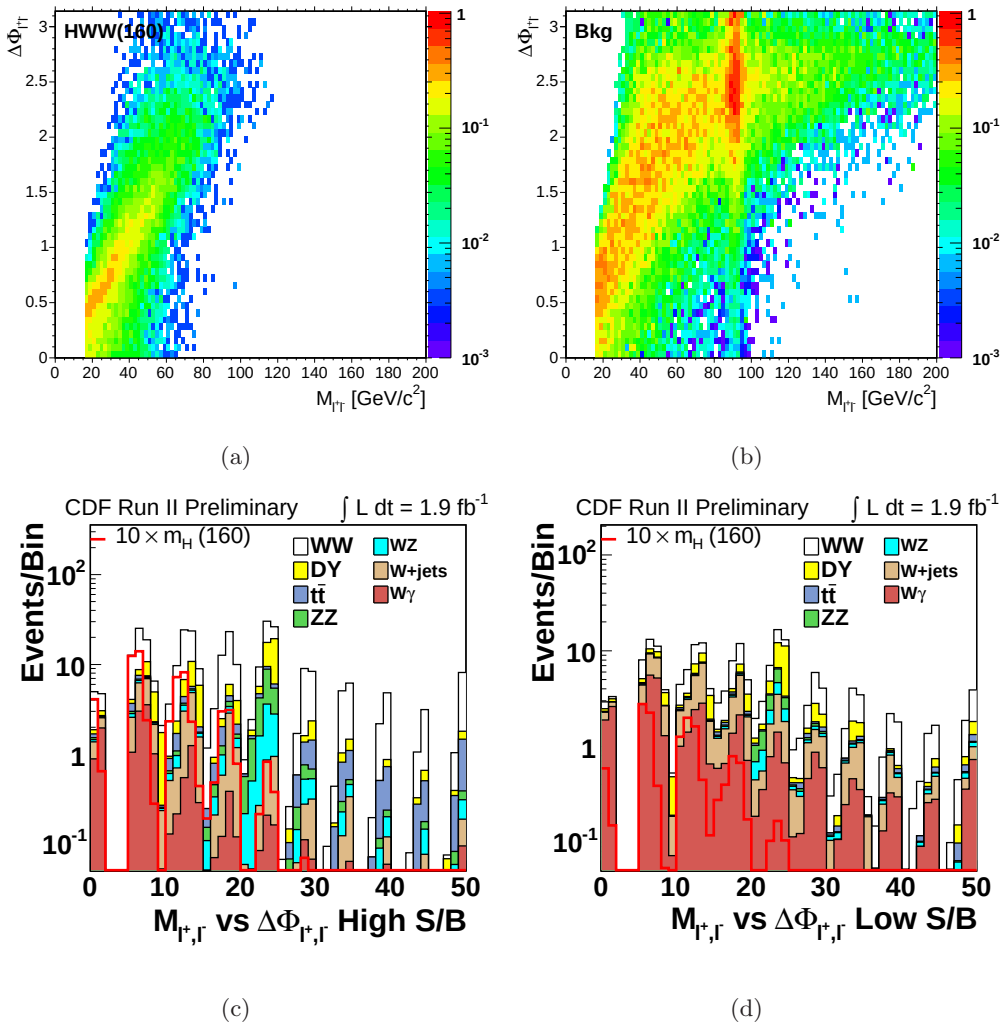


Figure 6.7: (a) and (b) are scatter plots of $\Delta\Phi(l^+l^-)$ v.s. $M(l^+, l^-)$ for Higgs(m_H=160) and other backgrounds. (c) and (d) are 1D projections of 2D distributions in high(low) S/B category. Each histogram is re-binned to 10×5 divisions.

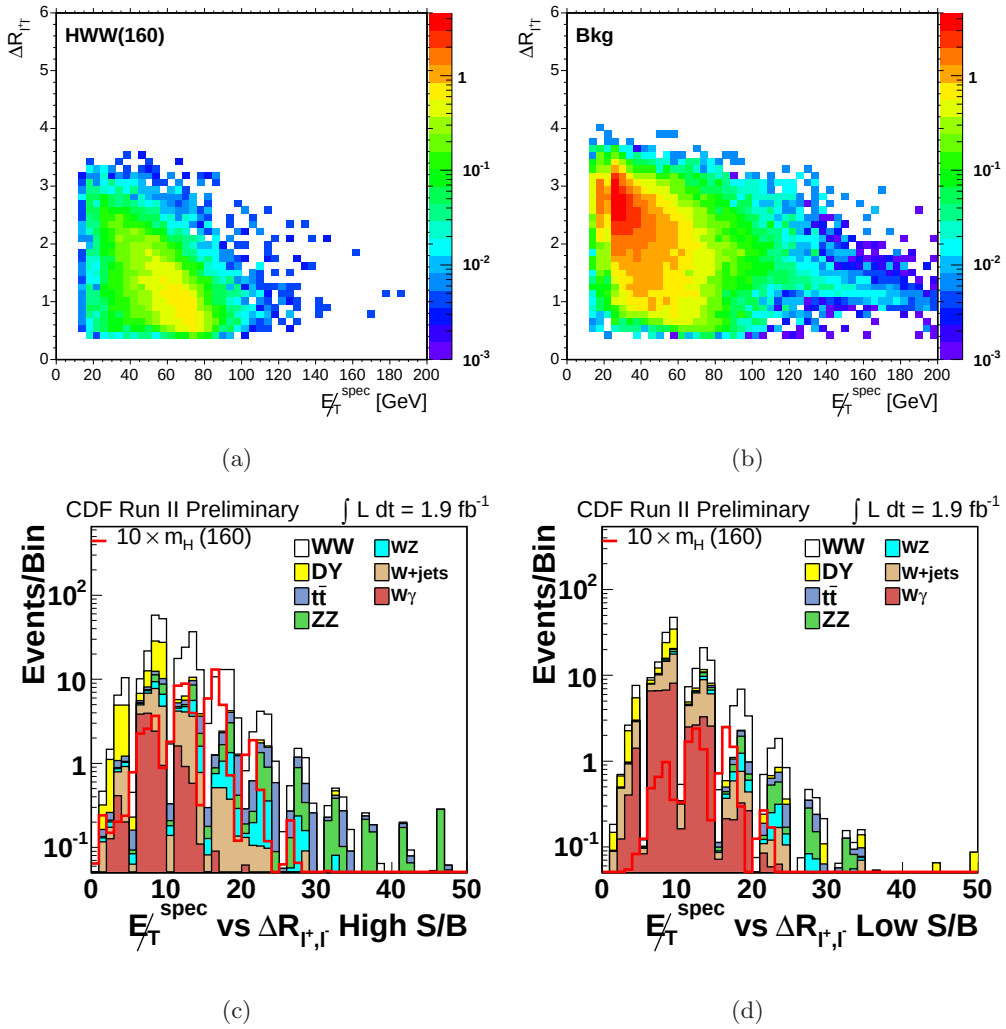


Figure 6.8: (a) and (b) are scatter plots of $\Delta R(l^+l^-)$ v.s. $\not\!\!E_T^{spec}$ for Higgs(m_H=160) and other backgrounds. (c) and (d) are 1D projections of 2D distributions in high(low) S/B category. Each histogram is re-binned to 10×5 divisions

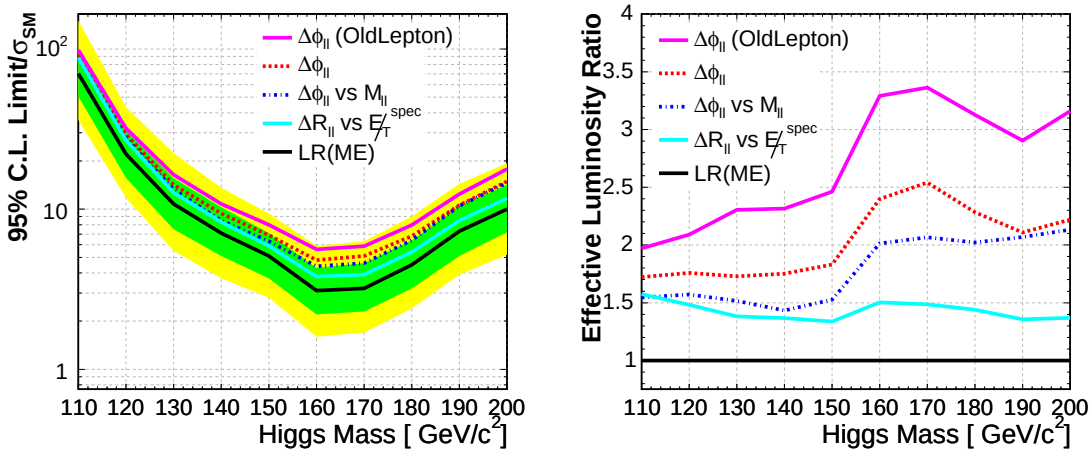


Figure 6.9: The sensitivity comparison of various discriminators. The right plot is normalized to the likelihood ratio discriminator with new lepton selections. In the case of $m_H = 160$ GeV/c^2, the effective accumlated luminosity gains are 40% and 140% due to adding new leptons and using LR discriminator, respectively.

Chapter 7

Systematics

The systematic uncertainties can affect normalizations and the shapes used in the limit calculation. We study the shape effects of each systematics and determine to include the shape variations into the fit or not. As a result, we only consider shape uncertainties for W+jets, which we expect them to be significant. The other uncertainties are treated as scale uncertainties when it is sufficient. The methods to evaluate various systematic errors are presented in the order of the importance of the contribution. We start from theoretical cross-section and then go through next-to-leading order acceptance, luminosity, $\not\!\!E_T$ modeling, trigger efficiency, lepton identification, parton density function, photon conversion and fake rates. An observation of 30% excess in DY control region with $16 < m_{l^+l^-} < 25\text{GeV/c}^2$ triggers studies of considering extra DY templates in the Base region. It turns out to be a negligible effect for limit calculation. A summary table and further discussions of the systematics are described in the last section.

7.1 Theoretical uncertainty

The production cross-section used for this thesis is calculated to next-to-leading order for most of the processes as shown in Table 5.2. For the WW, WZ, ZZ and $W\gamma$, we vary 10% of the next-to-leading order cross-section uncertainties from MCFM calculation [65]. The 15% $t\bar{t}$ cross-section uncertainty has been calculated at next-to-leading order in perturbation theory [66]. For Higgs cross-section, it has been calculated to next-to-next-to-leading order [64], we quote a 10% uncertainties.

7.2 Next-to-leading order uncertainty of acceptance

To assess the uncertainty of acceptance due to higher order effects, we generated WW in MC@NLO [60] (next-to-leading order monte carlo with Herwig parton shower) and went through full realistic cdfSim detector simulation. The differences of acceptances between WW in PYTHIA are used to determine the systematic errors. Since NLO calculations are the same physics for other background processes, the same systematics can be applied to all the rest of samples. The NLO acceptance uncertainties of Higgs are estimated by reweightig P_T of Higgs from PYTHIA to NLO calculation.

7.2.1 WW model

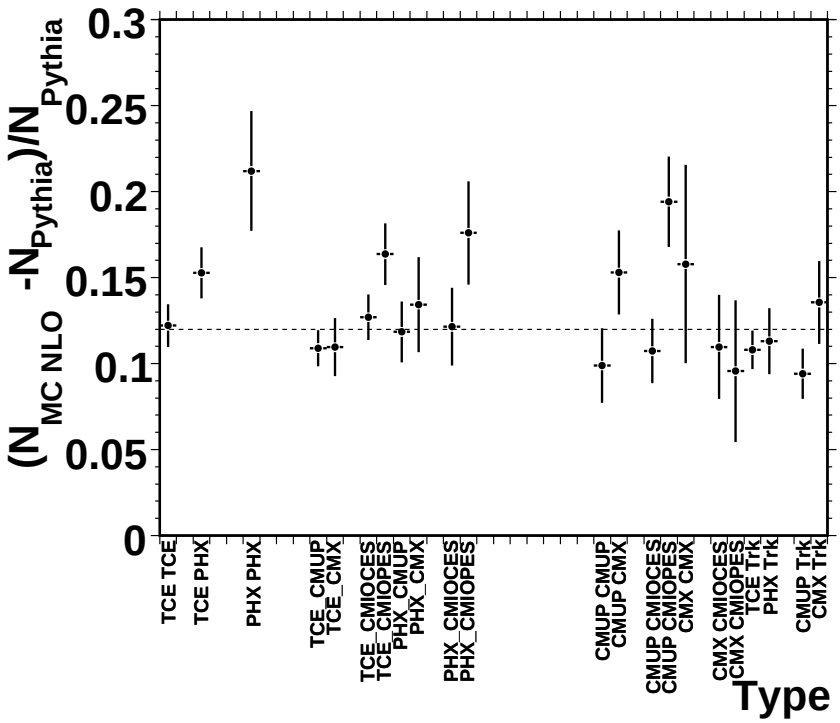


Figure 7.1: NLO effects on WW acceptance are evaluated by comparing MC@NLO to PYTHIA.

We used WW in MC@NLO as the base line templates for background modeling. The NLO acceptance uncertainties for WW are calculated by taking half of the % differences between PYTHIA and MC@NLO. The acceptance difference in % is equivalet

to the predicted yield difference in % since we normalize both WW Monte Carlos to the same production cross-section. The total acceptance of MC@NLO is about 12% larger but varies with different dilepton categories as seen in Figure 7.1.

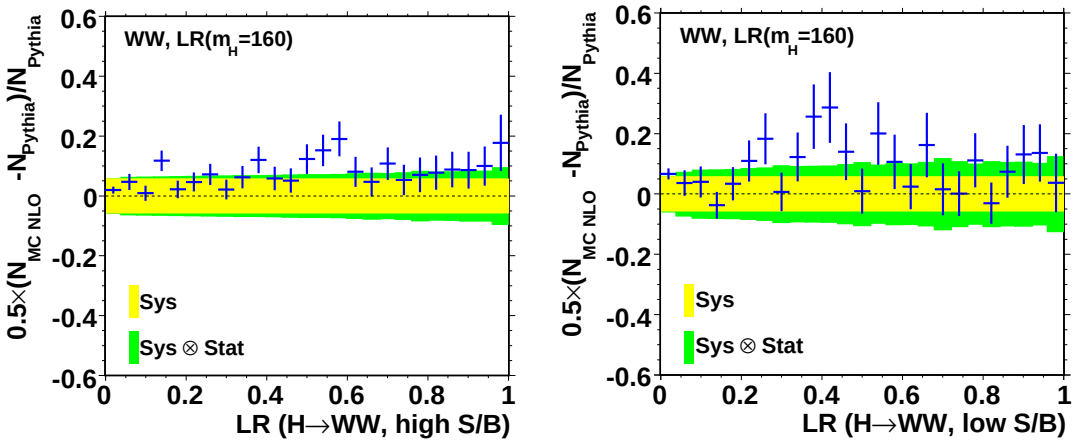


Figure 7.2: Likelihood ratio variations between MC@NLO and PYTHIA in %.

The effect of the NLO vs PYTHIA difference on the LR shapes can be seen in Figure 7.2. The differences between PYTHIA and MC@NLO in the LR variable are consistent with flat within the statistical uncertainty.. We decided not to apply a shape systematics at this point, but rather apply only the scale systematics listed in Table 7.6.

In order to gain better understandings of the sources of the additional acceptance, we studied yield difference in % for several kinematic variables which pass the other cut variables except the one of interest. The results are shown in Figure 7.3. The most significant excess of acceptance is in one jet bin and MC@NLO has relative 40% more acceptance. The other kinematics variables doesn't show significant difference except MC@NLO has shofter $\not\!\!E_T^{spec}$ spectrum and has slightly more back-to-back leptons in phi .

For the WZ, ZZ, $W\gamma$, and $t\bar{t}$ modes we include a 10% uncertainty set by the scale of this difference (not halved). For DY,we reduce this to 5% because DY has been well studied, and the PYTHIA MC has been tuned to reproduce several DY distributions (e.g. Z p_T), and is known to agree with data to much better than 10%.

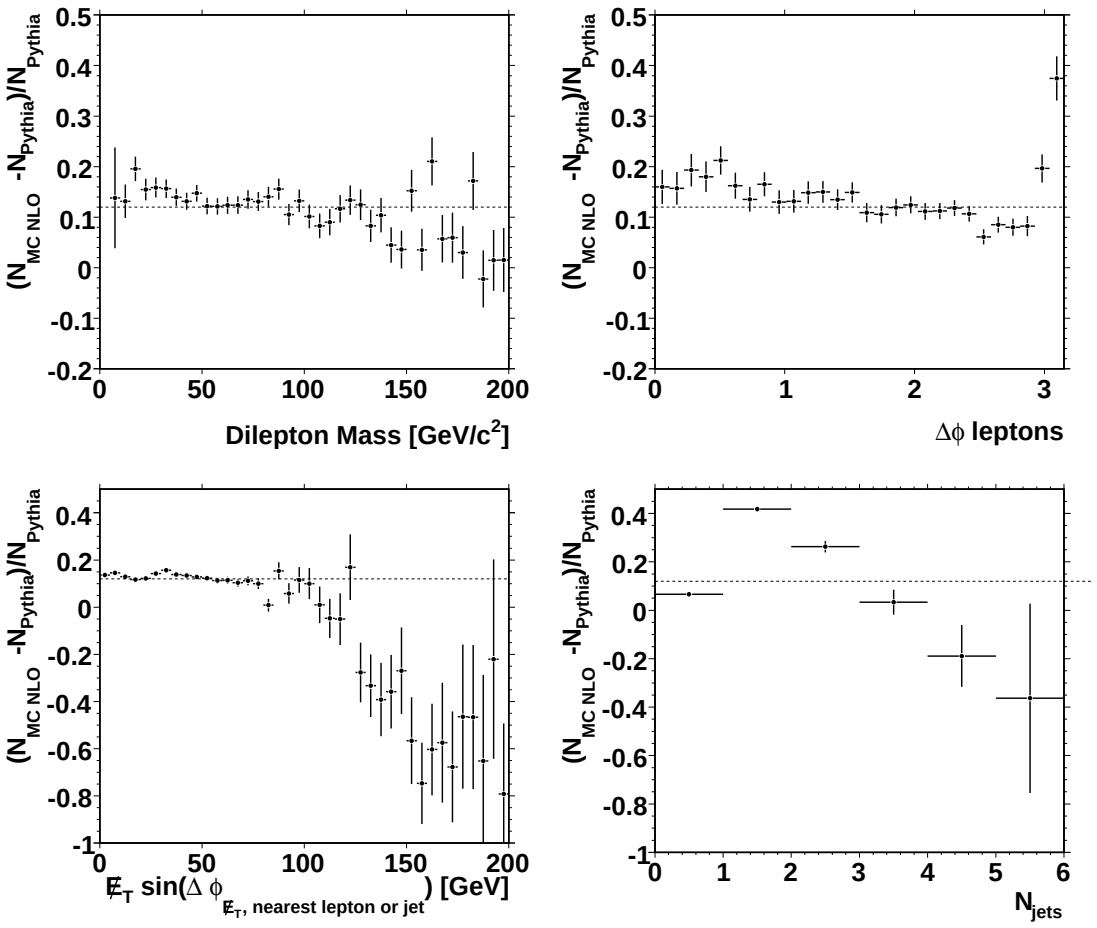


Figure 7.3: MC@NLO to Pythia comparisons in % for kinematic variables: $m(l^+l^-)$, $\phi(l^+l^-)$, $\not\!\!E_T^{spec}$ and number of jets.

7.2.2 $H \rightarrow WW$ model

The (N)NLO corrections to the Higgs cross-section are large, but because the Higgs is a pseudo-scalar, the only kinematic observables effected by higher-order corrections are the P_T and η; W spin correlations cannot be effected. As we do not use the jet energy distributions in this analysis, we are not sensitive to any NLO effects on those distributions. The vector sum of the hadronic recoil does enter the matrix element calculation through the $\not\!\!E_T$, but this is just the P_T of the produced Higgs. In order to assess the effect of the P_T modeling, we re-weight the Higgs sample based on P_T distribution in the Pythia MC. The NLO P_T distribution is calculated by the FEHiP program [74]. The P_T and η distributions of Higgs are shown in Figure 7.4. The resulting systematic

errors shown in Table 7.1 are of order 5%. We studied the LR distribution for different pT Higgs distributions, and find the effect negligible.

Table 7.1: Effect of re-weighting the Higgs p_T distribution from the PYTHIA to the FEHiP NLO calculation [74] in %.

Category	$\frac{(A\epsilon)_{\rm NLO}-(A\epsilon)_{\rm LO}}{(A\epsilon)_{\rm LO}}$ (%)									
Channel	110	120	130	140	150	160	170	180	190	200
$e\ e$	4.7	5.7	4.2	4.5	6.2	5.0	8.7	7.4	6.6	5.5
$e\ \mu$	2.0	4.5	4.6	4.6	5.8	4.6	6.0	5.1	5.3	5.5
$\mu\ \mu$	4.4	7.8	5.9	5.4	7.2	5.6	7.1	7.6	8.6	9.2
e trk	1.3	5.2	6.1	6.0	4.1	6.1	6.4	2.2	5.1	6.7
μ trk	4.2	3.5	4.4	5.7	7.2	3.4	7.5	6.7	7.5	6.9
Total	3.0	5.1	4.9	5.0	5.9	5.0	6.9	5.7	6.2	6.4

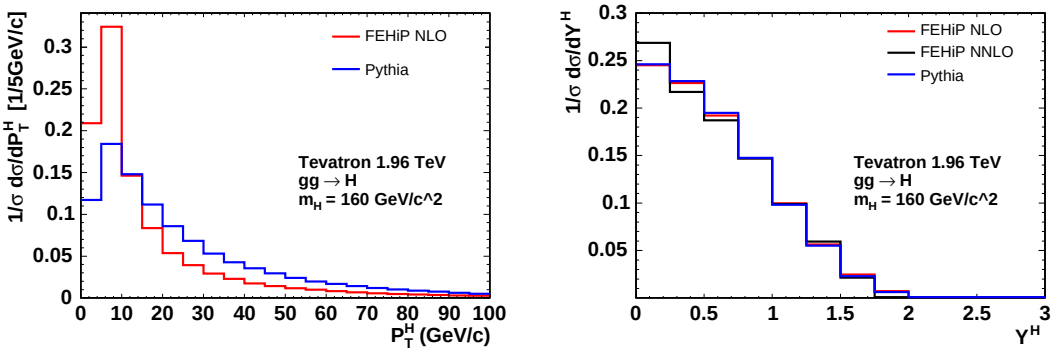


Figure 7.4: PYTHIA and FEHiP (N)NLO calculations of the Higgs P_T and rapidity distributions. The P_T distributions for NLO and PYTHIA are normalized to equal area. If we had plotted instead the cross-section normalized distributions then one would see that NLO adds only at low P_T to the PYTHIA distribution, i.e. at P_T of 40GeV and higher PYTHIA and NLO are identical.

7.3 $\not\!\!E_T$ Modeling

The uncertainty due to modeling of the $\not\!\!E_T$ and ΣE_T dependence is estimated in this section. The accuracy of the Monte Carlo in modeling the $\not\!\!E_T$ distribution can be assessed through the level of agreement in the Drell-Yan $\not\!\!E_T$ and ΣE_T distributions. It should be noted that the event with real $\not\!\!E_T$ is sensitive to the core of this distribution, while the event with fake $\not\!\!E_T$ is sensitive to the tail.

7.3.1 Events with real $\not\!\!E_T$

A quantitative estimate of the uncertainty is made by separately re-weighting the WW Monte Carlo events by the data over Monte Carlo ratio of $\not\!\!E_T$ and ΣE_T distributions. The weight used is

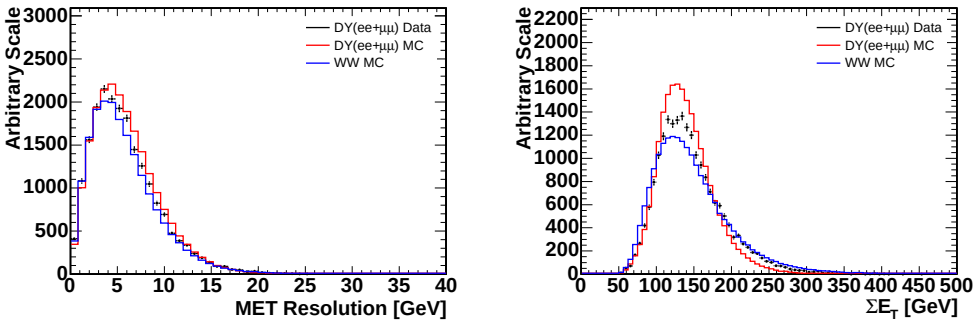
$$w = \frac{D^{DY}_{data}(|\vec{\not\!\!E}_{T\,reco}^{WW} - \vec{\not\!\!E}_{T\,reco}^{WW}|)}{D^{DY}_{data}(|\vec{\not\!\!E}_{T\,gen}^{WW} - \vec{\not\!\!E}_{T\,gen}^{WW}|)}, \tag{7.1}$$

where D_{data} and D_{MC} in Figure 7.5(a) is the distribution of the $\not\!\!E_T$ resolution for the WW Monte Carlo. The WW Monte Carlo resolution is re-weighted by the weight that would take the Drell-Yan Monte Carlo to the data. All cuts except $\not\!\!E_T$ are applied to these distributions. The result was found to be 1%. We apply a 1% systematic to all modes with real $\not\!\!E_T$

The weight function for ΣE_T systematic on the $\not\!\!E_T$ is defined:

$$w = \frac{D^{DY}_{data}(\Sigma E_T)}{D^{DY}_{data}(\Sigma E_T)}. \tag{7.2}$$

The distributions are shown in Figure 7.5(b) and the effect s less than 1%.



(a) (b)

Figure 7.5: Distributions used for (a) $\not\!\!E_T$ and (b) ΣE_T acceptance systematics.

7.3.2 Events with fake $\not\!\!E_T$ in Drell-Yan ee and $\mu\mu$ events

We use our $\not\!\!E_T$ control regions to limit the mismodeling of the Drell-Yan $\not\!\!E_T$ tail. Both regions are statistically consistent with the prediction, so we use the larger of the two statistical uncertainties (because they probe different effects using the smaller one might not be sufficient). The resulting numerical value is $1/\sqrt{29} = 0.186$ which is rounded to 20%.

7.4 Conversion rate and veto-efficiency in $W\gamma$

The $W\gamma$ mode enters the WW selection when the photon is reconstructed as an electron and we also have large $\not\!\!E_T$. This is usually due to a conversion with highly asymmetric E_Ts of each electron-positron pair. In the TCE electron, we explicitly veto conversions. Photon conversions requires two opposite sign COT tracks such that the $\Delta \cos\theta < 0.04$ and the distance between the two tracks at closest approach $D_{xy} < 0.2cm$. The partner COT track is passing p_T threshold > 0.7 GeV/c. The efficiency of this veto may not be well modeled in the Monte Carlo. An extra correction and its uncertainty is considered in this analysis.

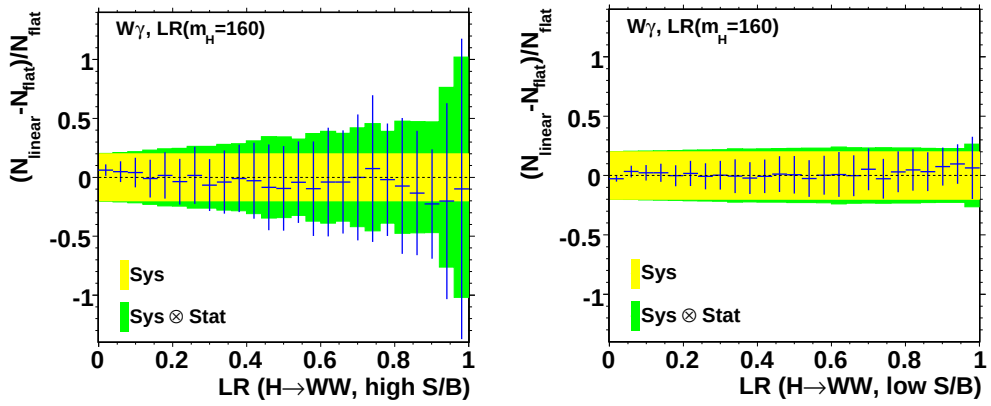


Figure 7.6: Shape function for Wγ by varying ± 1 sigma of the rescale constant for conversion veto efficiency.

The conversion tagging efficiency is measured by using tag-and-probe method from data and compared to Monte Carlo [75]. The scale factor has a functional form SF $= max[1, (1.979 - 0.0320 * E_T)]$. The LR template variation due to this scale factor is show in Figure 7.6. After inspection of Figure 7.6, we decided not to apply additional shape variation, but to use a flat correction and 20% scale uncertainty as systematics. This uncertainty dominates other uncertainties in $W\gamma$, such as NLO cross-section uncertainties.

7.5 PDF Systematics

We currently only us a normalization uncertainty for the PDF systematics using the standard CTEQ6M [76] uncertainties. Shape systematics are difficult for the

PDF because there are 40 variations that are prohibitively computationally intensive to implement in the limit calculator. We calculate the uncertainties using the standard CDF formulae,

$$\begin{aligned} \Delta A^{+} &= \sqrt{\sum_{i=1}^{i=20} \max\left(A_i^{+} - A_0, A_i^{-} - A_0, 0\right)^2} \\ \Delta A^{-} &= \sqrt{\sum_{i=1}^{i=20} \max\left(A_0 - A_i^{+}, A_0 - A_i^{-}, 0\right)^2} \end{aligned} \tag{7.3}$$

where i is the index of the CTEQ6M eigen value pair and A is the acceptance. The effect of the PDF uncertainty on the cross-section is taken into account separately in the cross-section/theoretical uncertainty. We add in quadrature the CTEQ6M variations with the difference between CTEQ5L and CTEQ6m because we generate our samples with CTEQ5L which may have larger uncertainties than CTEQ6M. The large numbers for $m_H = 110$ and $W(\rightarrow \tau\nu_\tau)\gamma$ in Table 7.2 are due to the low statistics in the reconstructed samples. We use common vales for all the $W\gamma$ and Higgs modes to avoid this problem. The numbers used are shown in the systematics summary Table 7.6.

7.6 Lepton Identification and Trigger Efficiency

The systematics of lepton identification efficiencies are estimated by varying the ID scale factors to $\pm 1\sigma$ deviation and recalculate the expected yields for each process. The statistical uncertainty on the scale factor is approximately 0.5% per lepton, so a systematic of 1% is roughly for total. The same procedure is used in the calculation of the systematics of trigger efficiencies. The result of the systematic due to trigger efficiency is about 2%. One additional systematic for the effect of the E_T thresholds of the electron triggers is included by counting 16 GeV instead of the nominal 20 GeV as triggerable.

$W\gamma$ is the background most affected by trigger systematics because it has conversion photon being identified as an electron. The systematics are dominated by forward electrons (PHX) where the trigger efficiency has a slow turn on curve of $\not\!\!E_T$. The fact that a W which puts an eletron in the plug will tend to also produce lower $\not\!\!E_T$ which enhance the effects. becomes important. The turn-on curve of the MET_PEM trigger is shown in Appendix A. Table 7.3 lists the variation of $W\gamma$ yields due to trigger efficiency

Table 7.2: The effect of the PDF variation on the acceptances in %.

Mode	ΔA^+	ΔA^-	Δ(CTEQ5L − CTEQ6M)	Total
WW	0.6	0.7	-1.8	1.9
WZ	2.5	0.7	-1.0	2.7
ZZ	0.1	2.7	-0.0	2.7
ttbar	1.8	1.2	-1.0	2.1
DYee	1.7	3.4	0.2	3.4
DYmm	1.7	4.0	-0.7	4.1
DYtt	3.3	5.2	0.1	5.2
wegamma	2.2	1.9	0.1	2.2
wmgamma	2.2	1.9	-0.1	2.2
wtgamma	5.5	20.1	4.6	20.6
HWW110	11.4	2.5	1.4	11.5
HWW120	2.5	1.6	-0.3	2.5
HWW130	0.9	2.4	-0.2	2.4
HWW140	0.2	1.7	-0.6	1.8
HWW150	0.5	0.8	-0.9	1.2
HWW160	1.3	1.8	-1.1	2.2
HWW170	0.7	1.8	-0.9	2.0
HWW180	0.7	2.0	-0.3	2.0
HWW190	1.7	1.7	-1.6	2.4
HWW200	1.2	1.9	-1.4	2.3

systematics.

Table 7.3: $W\gamma$ systematics variation due to trigger efficiency.

Category	Yields (stat)	Trig. sys. (%)
TCE TCE	5.1 ± 0.2	1
TCE PHX	14.6 ± 0.2	4
PHX PHX	7.4 ± 0.2	9
TCE CMUP	3.4 ± 0.2	4
TCE CMX	1.4 ± 0.1	2
TCE CMIOCES	1.3 ± 0.1	2
TCE CMIOPES	0.6 ± 0.1	2
PHX CMUP	7.3 ± 0.2	7
PHX CMX	4.5 ± 0.2	5
PHX CMIOCES	3.0 ± 0.2	21
PHX CMIOPES	2.3 ± 0.1	22
TCE Trk	2.8 ± 0.1	2
PHX Trk	3.7 ± 0.1	19
CMUP Trk	0.8 ± 0.1	6
CMX Trk	0.4 ± 0.1	3
Total	58.5 ± 0.6	7

7.7 W+jets modeling

We estimate the uncertainties on the fake background by varying the fake rates within their assigned uncertainties (shown by the solid light grey band in Figure 5.8(a), which are assigned based on the differences between the different jet samples.) The effects of these variations on the LR fitter are shown in Figure 7.7. The shape variation is included in the fit for final results.

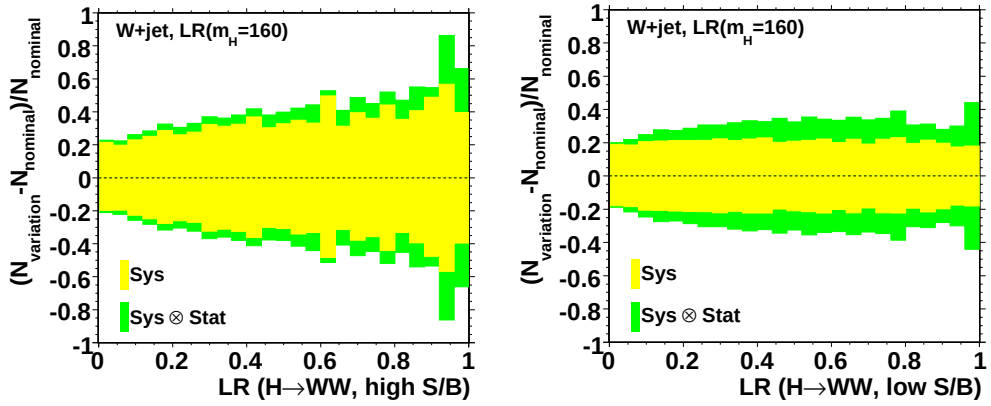


Figure 7.7: Shape function for Wjet by varying ±1 sigma of Fake Rate.

7.7.1 Special concern for $m_{ll} < 25$ GeV/c^2

In the Drell-Yan cross check region, we found 30% excess in the low dilepton mass between 16 GeV/c^2 and 25 GeV/c^2 as shown in Figure 7.8. The excess is 15% after requiring $Njets \leq 1$. We account for this excess by adding additional templates on top of the exisiting Drell-Yan templates. The Drell-Yan excess yields in the Base region are normalized to 0.181 and 0.099 for the high and low S/B categories, respectively.

The next question is to choose appropriate histograms as the LR template. Due to the low statistics of Drell-Yan Monte Carlo in this low mass range after Base selection, we construct LR shapes by loosening the $\not\!\!E_T^{spec}$ cut, e.g. $\not\!\!E_T^{spec} > 15GeV$(10 GeVfor e$\mu$). Figure 7.9 shows the likelihood ratio distributions of three different $\not\!\!E_T^{spec}$ cuts. The shape of no $\not\!\!E_T^{spec}$ cut is similar to the intermediate $\not\!\!E_T^{spec}$ cut.

Table 7.4 shows the expected limits including low mass Drell-Yan components with different shapes. The effect of excess is ignorable within 1000 pseudo-experiment uncertainties. Table 7.5 shows the expected limits of Higgs at mass 160 GeV/c^2 at

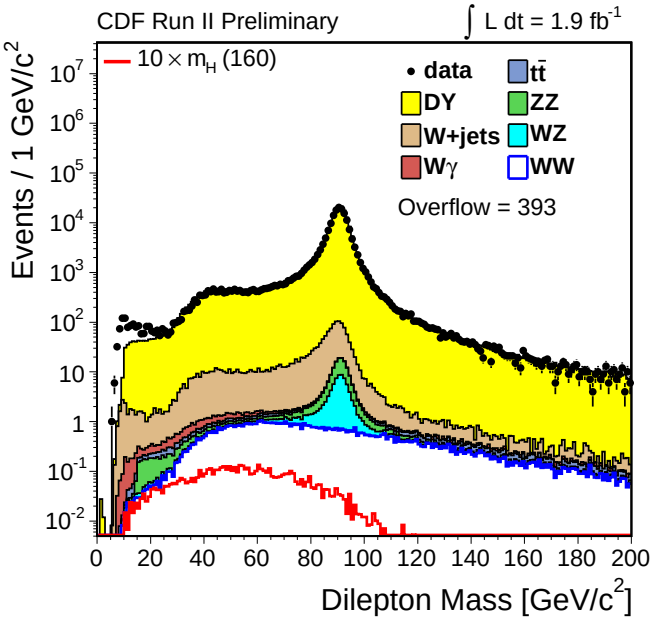


Figure 7.8: There are 30% excess observed in the $16 < m_{ll} < 25$ GeV/c^2 in the Drell-Yan region.

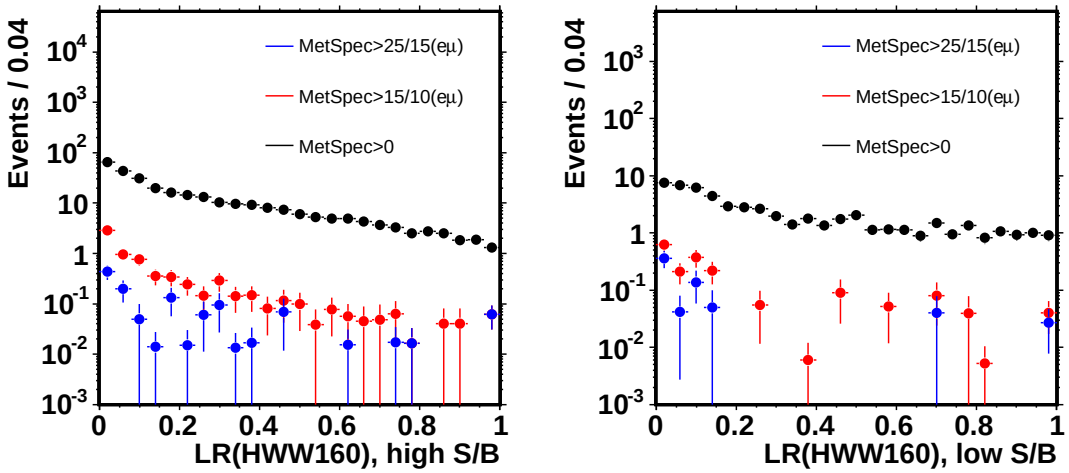


Figure 7.9: Likelihood Ratio distribution for Drell-Yan in the Base region between 16 GeV/c^2 and 25 GeV/c^2

different excess level. This test shows the effect of mis-modeling at low mass region is noticeable only when the excess yield is a factor of 10 larger.

Table 7.4: Expected limit for Higgs by adding extra Drell-Yan component with different shapes.

Higgs Mass (GeV/c^2)	100	120	130	140	150	160	170	180	190	200
Nominal	71.35	21.79	11.07	7.04	5.10	3.06	3.20	4.50	7.26	9.91
Higgs Shape	70.68	22.04	10.85	7.09	5.05	3.04	3.17	4.46	7.23	9.82
$\not\!\!E_T^{spec} > 15/10(e\mu)$	70.53	21.83	10.90	7.18	5.17	3.04	3.22	4.56	7.37	9.88
No $\not\!\!E_T^{spec}$	69.46	22.17	10.86	7.14	4.93	2.99	3.17	4.49	7.22	10.17

Table 7.5: Expected limits for Higgs with mass 160 GeV/c^2 at different excess ratio. The Higgs shape is used as low dilepton mass DY to study the scaling.

Excess (%)	0	15	56	200	1000
Expected limit	3.06	3.04	2.89	2.99	3.55

7.8 Summary

The resulting scale systematics applied are listed in Table 7.6. The linear correlation of each systematics sources are considered and evaluated in the "Sum" column. The formula used for calculation is

$$\text{Sum} = \sum \text{N}_\text{i}\delta_\text{i}, \tag{7.4}$$

where N_i and δ_i are the expected yields and systematic errors for each background process. The major systematic error is next-to-leading-order cross-section uncertainty 8%. It is mainly coming from WW backgrounds which occupy half of the total backgrounds. The second largest systematics is next-to-leading-order acceptance 6%. The improvement of this analysis could be done by measuring the ratio of production of H over Standard Model WW. An update of the measurement of WW production cross-section could then be used to re-scale the absolute production cross-section of Higgs.

The squared root quadratic sum of all background sources are calculated in each column as the systematics fluctuation for each process. As a result, the total systematics error for $W\gamma$ is the largest and mainly due to uncertainty on the rate of conversions.

Table 7.7 shows the impact of different systematic errors on the expected Higgs cross-section limit. We take one type of systematics at a time, and quote the relative change in the limit with respect to applying no systematics. All numbers in the table are thus in terms of % change of the limit due to applying systematics. 10,000 pseudo-experiments are used to calculate each limits. The dominant sources of systematics are NLO cross-section and NLO acceptance uncertainties. The $\not\!\!E_T$ modeling has larger

Table 7.6: Summary of the Systemtics in % . The W+jets systematics for the high and low S/B subsamples separately are 19.1 and 23.3%. The high S/B subsample has a larger uncertainty because the fake rate determination for these has lower statistics.

	WW	WZ	ZZ	$t\bar{t}$	DY	$W\gamma$	W+jets	Sum	Higgs
Cross-section	10.0	10.0	10.0	15.0	5.0	10.0	-	8.0	-
NLO Acceptance	5.5	10.0	10.0	10.0	5.0	10.0	-	5.6	10.0
Luminosity	6.0	6.0	6.0	6.0	6.0	6.0	-	5.2	6.0
$\not\!\!E_T$ Modeling	1.0	1.0	1.0	1.0	20.0	1.0	-	3.9	1.0
Trigger Eff	2.1	2.1	2.1	2.0	3.4	7.0	-	2.6	3.3
LepID	1.5	1.4	1.3	1.5	1.5	1.2	-	1.3	1.5
PDF Uncertainty	1.9	2.7	2.7	2.1	4.1	2.2	-	2.1	2.2
Conversions	-	-	-	-	-	20.0	-	2.3	-
Jet Fake	-	-	-	-	-	-	22.5	3.0	-
Total	13.3	10.2	15.8	19.3	22.7	26.3	22.5	12.9	12.4

Table 7.7: The systematics variation of Higgs production cross section limits due to different sources.

Type	110	120	130	140	150	160	170	180	190	200
Cross-Section	10	10	15	16	17	10	12	13	17	20
NLO Acceptance	12	12	16	17	15	10	11	13	18	20
Luminosity	10	10	13	15	13	9	9	11	15	16
$\not\!\!E_T$ Modeling	9	7	7	7	6	2	3	4	5	5
Trigger Eff.	5	4	6	5	5	3	3	3	2	4
LepID	2	1	3	1	2	1	1	1	2	1
PDF	4	3	3	4	4	1	3	3	3	3
Conversion	3	1	3	1	1	0	1	-1	-0	-0
Jet Fake	3	3	4	3	2	1	1	1	0	0
Total	17	18	22	23	23	14	17	20	25	28

effects for $m_H < 150$ GeV/c^2 since dilepton invariant mass is small and contaminated by low mass Drell-Yan.

Chapter 8

$H \rightarrow WW^*$ Production Cross-Section Limits

In this chapter, we present the results of Higgs search with 1.9 fb^{-1} data. Section 8.1 shows the proof of principle of the implementations of Matrix Element method in this thesis. We re-define the numerator of likelihood ratio variable to be one of the background event probabilities and perform statistical tests over data. Section 8.2 analyzes the uncertainties of likelihood ratio variable calculation which using Importance Sampling technique to calculate event probabilities. Section 8.3 reports the expected and observed limits with/without systematics. Section 8.4 further discusses how the results depend on the different backgrounds. We study the effect of expected 95% C.L. cross-section limits by changing various background contributions or systematics one at a time. Section 8.5 gives a final summary.

The result was unblinded twice, first with 1.1 fb^{-1} and a second time with the full 1.9 fb^{-1} which this thesis is based on. The two results were presented in March Moriond 2007 conference [77] and August Lepton-Photon 2007 conference [7], respectively. All the tools and analysis strategies developed for this thesis continue to be in use at CDF. More data will thus be added to this analysis as the data becomes available. Section 8.3.2 shows the results with 1.9 and 3 fb^{-1} which is used for journal publication [10]. This analysis is expected to be updated once or twice a year until CDF has stopped data taking in 2009 or 2010.

8.1 Matrix Element Tests

To demonstrate that the implementation of phase space integration works correctly, a series of tests were done before un-blinding the Higgs likelihood ratio. In analogy to the likelihood ratio discriminator variable we use for separating higgs events from background, we can define discriminants for each of the background hypotheses. Figure 8.1 shows the distributions of these LR variables for the three dominant background distributions with the data overlayed. The overall agreement between data and expectations gives us confidence that the background models accurately describe the distribution of the data across phase space.

Using these background likelihood ratios described above, we can use the limit calculation procedure to calculate limits for the WW, W+jets, and $W\gamma$ just as we do for the Higgs signal, but treating one of these backgrounds as signal.

We thereby measure the WW, W+jet, and $W\gamma$ yields normalized to their prediction. In Table 8.1, we report the bounds on the ratio of observed to expected. These were obtained by integrating the distributions returned by the limit calculator up to the 5%($\approx -2\sigma$), 16%($\approx -1\sigma$),84%($\approx +1\sigma$), and 95%($\approx +2\sigma$) points. For example the region [0.75,1.04], contains the central 68% of the probability for the WW yields. Seeing that the value 1.0 is within this range implies that our measurement of the WW cross-section is within 1σ of the NLO theory prediction.

We conclude that we can use the appropriate LR distributions to measure the three dominant backgrounds, and that those measurements are consistent with expectations within their statistical and systematic errors.

Table 8.1: Results of extracting bounds for WW, W+jets, and $W\gamma$ using likelihood ratios with the respective mode treated as the signal. The numbers are as a fraction of their expectation.

Mode	5%($\approx -2\sigma$)	16%($\approx -1\sigma$)	84%($\approx +1\sigma$)	95%($\approx +2\sigma$)
WW	0.67	0.75	1.04	1.15
W+jets	0.57	0.82	1.67	2.00
$W\gamma$	1.08	1.30	2.36	3.05

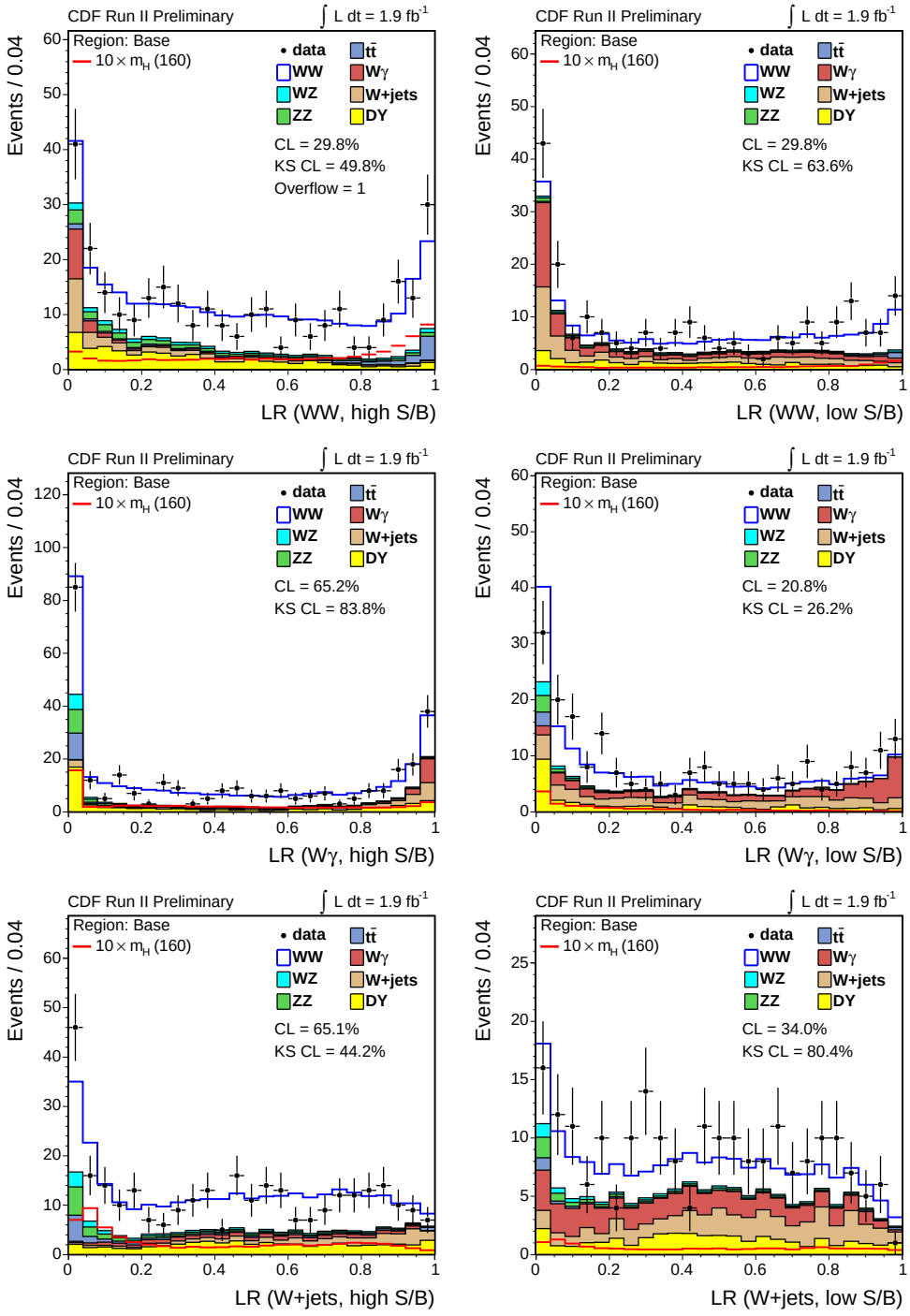


Figure 8.1: Likelihood ratio distributions for WW, $W\gamma$ and W+jets as signal hypothesis in Base region

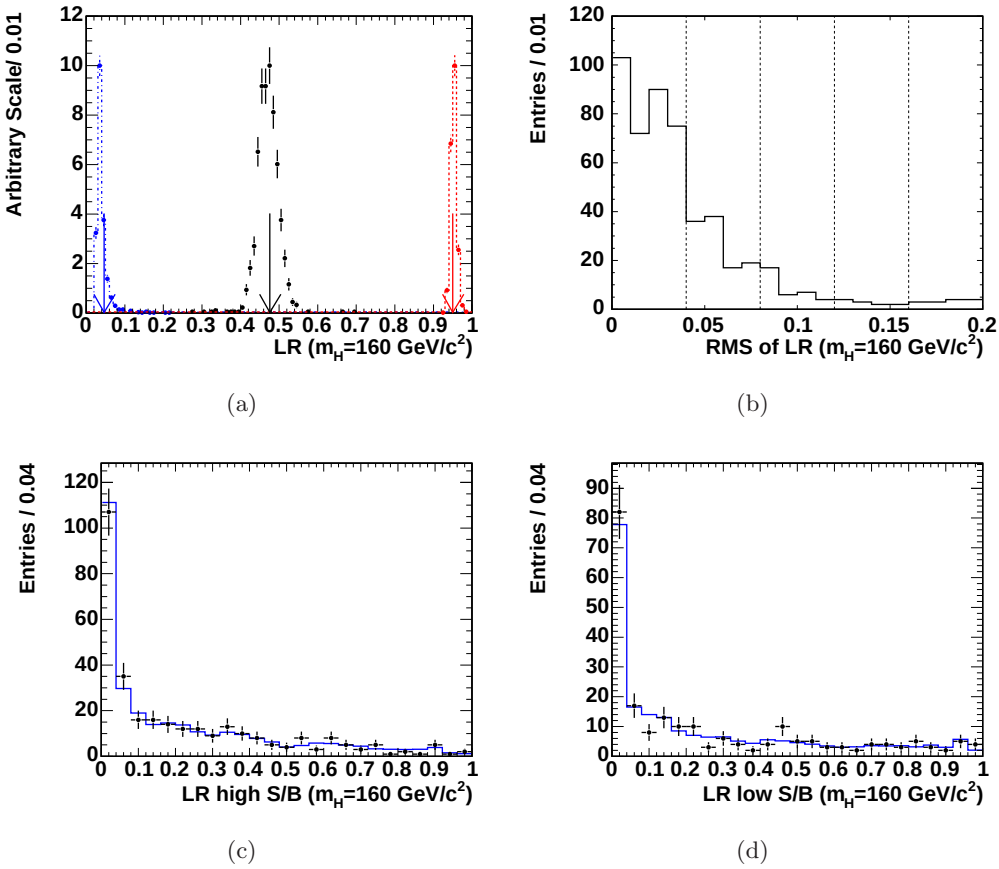


Figure 8.2: Likelihood Ratio distribution for Higgs (m_H = 160 GeV/c^2). (a) The distribution of LR with different random seeds for three events (b) The root mean square distribution of LR values for 522 candidate events (5 overflow entries). (c)(d) The averaged LR templates for high(low) S/B channels compared to the single seed events used in the final results.

8.2 Uncertainties of Event Probability Calculation

Importance sampling method is officially used to calculate the event probabilities in this analysis. Due to the limited computing resources, we used seed=1 to caculate event probabilities for each candidate event. There are 40K importance sampling points used for each event probability. The finite number of sampling points bring a potential problem: the random seed initialization affects the score of each event probability and therefor changes likelihood ratio score. This in-accuracy turns out not to be a problem since we are not doing unbinned likelihood fitting but a binned histogram fitting. As long as the shapes of data histograms are not sensitive to random seeds, we will get

Table 8.2: Base region Yields divided into the low and high S/B categories described in the text.

Category	WW	WZ	ZZ	$t\bar{t}$	DY	$W\gamma$	W+jets	Total	Data
HighSoverB	185.4	11.6	15.2	13.8	53.3	15.9	25.6	320.7 ± 26.2	305
LowSoverB	65.7	5.3	5.7	3.1	28.9	42.6	41.1	192.3 ± 17.3	217

	Higgs Mass (GeV/c^2)									
Category	110	120	130	140	150	160	170	180	190	200
HighSoverB	0.3	1.1	2.4	3.9	5.1	6.3	6.2	5.0	3.6	2.8
LowSoverB	0.1	0.3	0.6	0.9	1.3	1.4	1.4	1.2	0.9	0.7

consistent statistical interpretations.

Figure 8.2(a) shows LR scores of three candidate events. Each has 1000 different random seeds and the arrows indidate the LR scores with seed=1. These three candidate events have root mean square (RMS) less than 0.01 and have high, medium and low scores, respectively. Figure 8.2(b) shows the RMS distributions of observed candidates with 1.9 fb^{-1} data. 67% of the 522 candidates have RMS smaller than 0.04 which is the bin size of LR histograms. The data histograms of averaged LR scores are the solid histograms in Figure 8.2(c)(d). The cross marks are data histograms with seed=1 and used to set final limits. The consistent shape distributions between single seed histogram and 1000k-seed-averaged histograms demonstrate that the results are not sensitivity to random seeds.

8.3 Results

In Section 6.3, we divide the total sample into two categories based on signal to background ratio. The predicted and observed yields of high S/B and low S/B channel in $1.9fb^{-1}$ are shown in Table 8.2. Figure 8.3 shows the likelihood ratio discriminator distributions of these two categories.

There are no significant excess events observed in the data and therefore we set a 95% C.L. production cross-section limit based on the Baysians approach described in Section 6.4.1. Table 8.3 shows the resulting medium expected and observed limits as a function of m_H without systematics variation. Table 8.4 shows the corresponding limits after taking systematics into account. Figure 8.4 visually compares the expected and actual limits relative to the Standard Model Higgs production cross-section. The

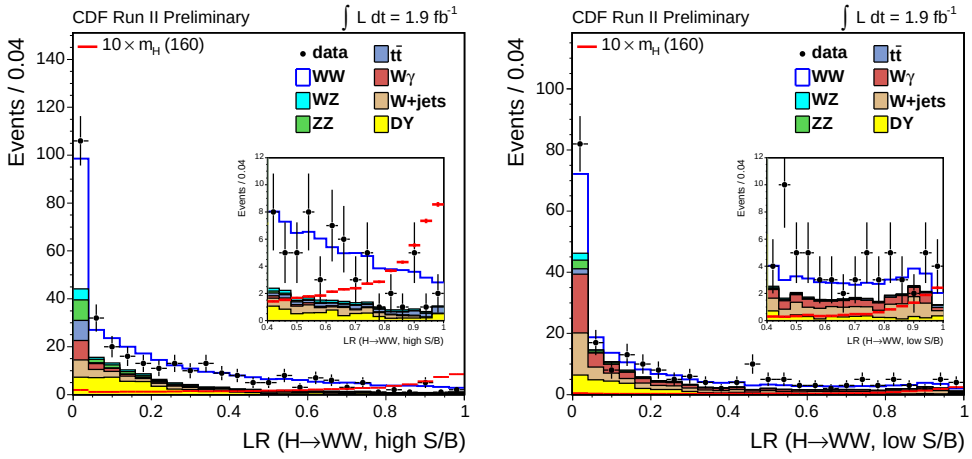


Figure 8.3: Likelihood ratio distribution for $m_H = 160$ GeV/c^2 in the Base region.

Table 8.3: Expected and observed limits for Higgs at 1.9 fb^{-1} **WITHOUT SYTEMATICS**.

$M_H(GeV/c^2)$	110	120	130	140	150	160	170	180	190	200
$\sigma_{SM}(HWW)(pb)$	0.057	0.134	0.230	0.312	0.358	0.388	0.344	0.278	0.194	0.155
$median(pb)$	3.4	2.6	2.1	1.8	1.5	1.0	1.0	1.1	1.1	1.2
$Observed(pb)$	4.3	2.6	1.5	1.3	1.1	0.7	0.8	0.8	1.2	1.5
$+2\sigma/\sigma_{SM}$	115.1	36.1	17.3	11.1	8.1	5.0	5.4	7.5	11.1	14.9
$+1\sigma/\sigma_{SM}$	84.2	27.0	13.1	8.0	5.8	3.8	3.9	5.5	8.2	10.8
$median/\sigma_{SM}$	59.5	19.1	9.3	5.7	4.1	2.7	2.8	3.8	5.7	7.6
$-1\sigma/\sigma_{SM}$	43.3	13.4	6.7	4.1	3.0	1.9	2.0	2.6	4.1	5.5
$-2\sigma/\sigma_{SM}$	31.7	9.9	5.2	3.2	2.4	1.4	1.5	2.1	3.1	4.1
$Observed/\sigma_{SM}$	75.9	19.3	6.6	4.3	3.0	1.9	2.3	2.8	6.0	9.4

Table 8.4: Expected and observed limits for Higgs production at 1.9 fb^{-1} with Systematics (the ratios include the uncertainty on σ_{SM}).

$M_H(GeV/c^2)$	110	120	130	140	150	160	170	180	190	200
$\sigma_{SM}(HWW)(pb)$	0.057	0.134	0.230	0.312	0.358	0.388	0.344	0.278	0.194	0.155
$median(pb)$	3.9	2.9	2.5	2.2	1.8	1.2	1.1	1.3	1.4	1.6
$Observed(pb)$	4.7	2.8	1.6	1.5	1.1	0.8	0.8	0.8	1.4	1.8
$+2\sigma/\sigma_{SM}$	138.4	42.1	21.7	13.2	10.0	6.0	6.3	9.0	14.3	19.3
$+1\sigma/\sigma_{SM}$	97.6	31.4	15.4	9.9	7.2	4.3	4.8	6.5	9.9	13.8
$median/\sigma_{SM}$	68.8	21.9	10.7	7.0	5.0	3.1	3.2	4.7	7.0	10.0
$-1\sigma/\sigma_{SM}$	49.1	15.4	7.7	4.8	3.6	2.2	2.2	3.2	5.1	7.1
$-2\sigma/\sigma_{SM}$	36.5	11.2	5.3	3.5	2.6	1.6	1.7	2.3	3.8	5.1
$Observed/\sigma_{SM}$	81.9	20.6	7.0	4.7	3.2	2.0	2.4	3.0	7.0	11.7

observed limit shows no sign of a signal and is consistent with the expected limit.

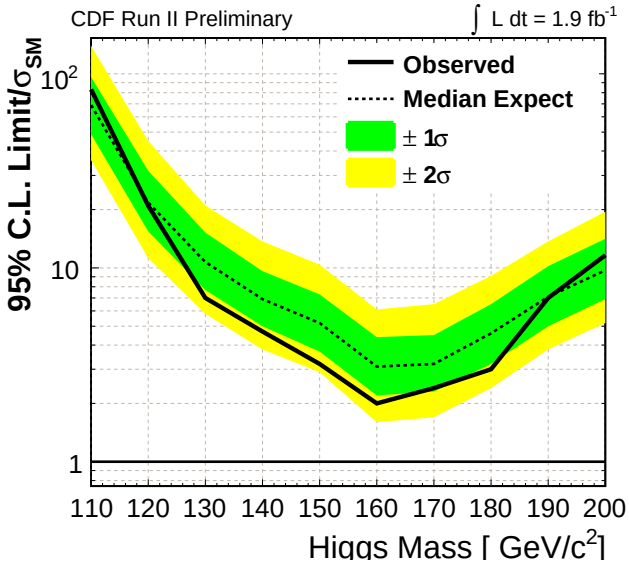
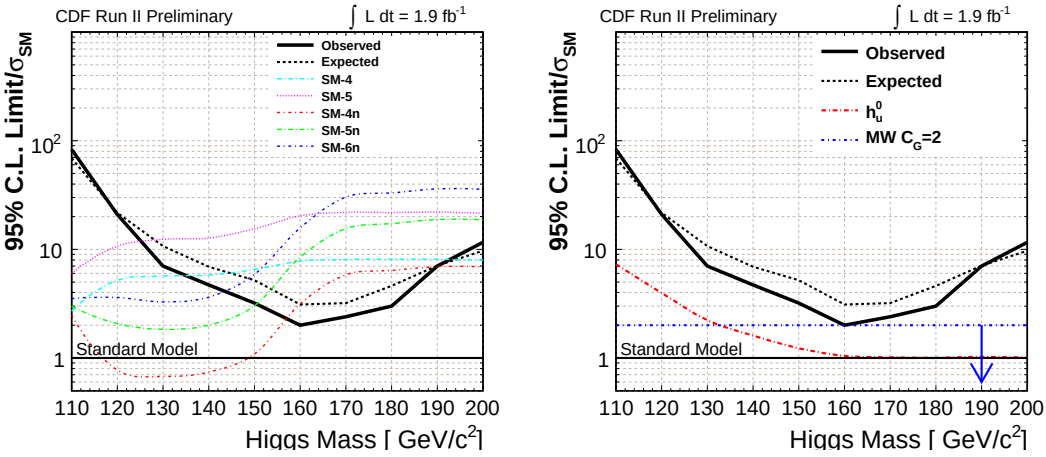


Figure 8.4: The ratio of 95% C.L. limit of $H \rightarrow WW$ production cross-section to SM cross-section prediction with 1.9 fb^{-1} data.

8.3.1 New Physics Model

In Section 2.2.4, we describe three new physics models which can enhance the production rates of $H \rightarrow WW^*$. In this section, we compare the observed and expected limits to each new physics model. Figure 8.5(a) shows the comparisons of observed limits to various extra generation models. We rule out $N > 4$ generation models (SM-N) for $m_H > 135$ GeV/c^2. In the extra generation model [27, 28] with one heavy neutrinos $m = 50$ GeV/c^2, we rule out $N > 4$ generation model (SM-Nn) for $m_H > 160$ GeV/c^2. Figure 8.5(b) compares observed limits to a up-type family coupling only Two Higgs Doublet model [29] and Manohar-Wise model [30]. In the up-type family coupling only Two Higgs Doublet model, there is no $Hb\bar{b}$ nor $H\tau\tau$ coupling such that the WW branching is enhanced. However, we are still not sensitive to this specific Two Higgs Doublet model with 1.9 fb^{-1} dataset. In Manohar-Wise model, Higgs can couple to gluon fields directly via 6 dimensional operators and the coupling strength is controlled by parameter C_G. The $H \rightarrow WW$ production can be any number above 0 with ignorable m_H dependence. With current results, we can constraint $C_G < 2$ at Tevatron.



(a) (b)

Figure 8.5: Expected and observed limits of $H \rightarrow WW$ production at 1.9 fb^{-1}. The results are compared to N generation models, up-type family coupling only Two Higgs Doublet model and Manohar-Wise model.

8.3.2 Limits with 3 fb^{-1} Data

After the completion of this thesis in June 2008, we have added more data and updated this analysis with 3 fb^{-1}. We expect 11.6 signal events (m_H = 160 GeV/c^2), 768±91 background events and observed 779 events. More details are shown in Table 8.5. We report 95% C.L. production cross-section limits as a function of m_H since there are no significant excess at every test point. The expected and observed cross-section limits are shown in Table 8.6. Figure 8.6 plots the ratio of 95% C.L. cross-section limit to the standard model NNLL cross-section calculation as a function of m_H. The expected limit of m_H = 160 GeV/c^2 production cross-section 0.9(1.2) pb at 95% C.L. which is 2.2(3) times the standard model cross-section at 3(1.9) fb^{-1} data. The 25% improvement of limit is purely coming from using more data which scales with square root of luminosity ratio. The CDF and D0 combined results have been presented at ICHEP 2008 [78]. After including 2 jet bin and consider all the Higgs production channel (WH, ZH, Weak Boson Fusion) process, Tevatron rules out the standard model Higgs production with 170GeV/c^2 at 95% C.L [78].

Table 8.5: Expected and observed yields in the Base region with 3 fb^{-1}.

Category	WW	WZ	ZZ	$t\bar{t}$	DY	$W\gamma$	W+jets	Total($\pm$ stat.)	Data
HighSoverB	263.3	17.2	16.1	20.9	86.5	23.8	44.6	472.4 $\pm$ 3.5	465
LowSoverB	92.2	7.7	5.8	4.6	51.7	66.7	65.4	294.1 $\pm$ 2.9	314
Sum	355.5	24.9	21.8	25.5	138.2	90.5	110.0	766.5 $\pm$ 4.5	779

	Higgs Mass (GeV)									
Category	110	120	130	140	150	160	170	180	190	200
HighSoverB	0.4	1.5	3.5	5.6	7.5	9.5	9.0	7.3	5.1	4.1
LowSoverB	0.1	0.4	0.8	1.4	1.8	2.1	2.0	1.7	1.2	1.0
Sum	0.5	1.9	4.3	7.0	9.3	11.6	11.0	9.0	6.4	5.1

Table 8.6: The 95% C.L limit of $H \rightarrow WW$ production cross-section as a function of m_H with 2.4 fb^{-1}.

	110	120	130	140	150	160	170	180	190	200
$-2\sigma/\sigma_{SM}$	37.89	11.26	5.50	3.48	2.52	1.50	1.51	2.12	3.39	4.74
$-1\sigma/\sigma_{SM}$	50.27	15.21	7.29	4.72	3.34	1.91	1.99	2.95	4.54	6.16
Median/σ_{SM}	**71.42**	**21.63**	**10.43**	**6.62**	**4.65**	**2.67**	**2.82**	**4.20**	**6.45**	**8.71**
$+1\sigma/\sigma_{SM}$	102.19	30.60	14.78	9.19	6.45	3.83	3.96	5.94	9.28	12.69
$+2\sigma/\sigma_{SM}$	135.63	42.24	20.16	12.70	9.27	5.40	5.55	7.64	12.49	17.13
Observed/σ_{SM}	**64.78**	**15.90**	**5.50**	**3.60**	**2.74**	**1.68**	**1.85**	**2.92**	**5.61**	**11.41**

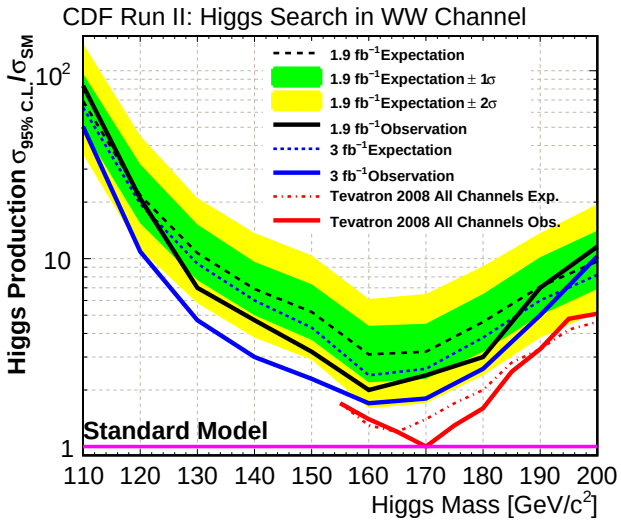


Figure 8.6: The ratio of 95% C.L. limit of $H \rightarrow WW$ production cross-section to SM cross-section prediction with 3 fb^{-1} data. The CDF and D0 combined limits (Tevatron results) with additional Higgs production channels and multiple jet bins have been ruled out the 170 GeV/c^2 Higgs at 95% C.L [78].

8.4 Discussions

To understand the importance of each background separately, we perform two additional tests. First, we vary systematics on one and only one background one at a

Table 8.7: The percentage difference of production cross-section limits between one backgroud only systematics variation and no systematics variation.

Type	110	120	130	140	150	160	170	180	190	200
WW	6	8	14	15	15	9	11	12	15	17
WZ	1	0	1	1	0	0	-0	-0	0	-0
ZZ	0	0	1	1	0	0	-0	-1	-0	-0
$t\bar{t}$	1	0	1	1	1	0	1	1	1	1
DY	8	6	6	5	4	1	2	2	3	3
$W\gamma$	4	2	3	2	1	1	2	0	0	0
Wjet	3	3	4	3	2	1	1	1	0	0
Higgs	5	5	4	4	4	4	5	4	4	4

time. Second, we eliminate a given background from template fitting one at a time. We generate 10000 pseudo-experiments to keep the statistical uncertainties of expected limits at 1% level.

To estimate how important the variation of each background is, we vary systematics on one and only one background one at a time. Table 8.7 shows the percentage differences of limits between one background only systematics variation and no systematics variation. The results are compared to the limits without systematics and shown in Table 8.7. WW-only variation affects limits by 6%$\sim$17% for different m_H. It has the largest effects over other backgrounds for $m_H > 120$ GeV/c^2. For $m_H < 160$ GeV/c^2, DY-only variation gradually competing with WW-only variation and has the largest effects for $m_H < 110$ GeV/c^2. $W\gamma$-only and Wjet-only has $\sim 3\%$ effects for $m_H < 160$ GeV/c^2 and $\sim 1\%$ for $m_H > 160$ GeV/c^2. The variation on Higgs systematics have a consistent $\sim 4\%$ effects.

To estimate the relative importance of the modeling of the background rates, we eliminate the background processes one at a time and recompute the production cross-section limits. Table 8.8 shows the percentage differences of limits between eliminating one background and including all backgrounds. WW plays the most important role of the whole Higgs mass range with the limits varying by 23% to 45%. For $m_H < 110$ GeV/c^2, DY plays an important role with the limit changing by 12% which is half of the effect of WW. WW is the most important background. All others are irrelevant in comparison. WZ and ZZ are irrelevant for all m_H sensitivity. $t\bar{t}$ is irrelevant expect for very high m_H where it becomes the second most important background. The importance of DY, $W\gamma$ and W+jet increases with decreasing m_H.

Table 8.8: The percentage difference of production cross-section limits between eliminating one background and all backgrounds.

Type	110	120	130	140	150	160	170	180	190	200
WW	-23	-32	-38	-43	-45	-42	-43	-43	-42	-41
WZ	0	-2	-0	0	-1	0	-1	-2	-3	-2
ZZ	1	0	1	1	0	1	-1	-1	-1	-1
$t\bar{t}$	-1	-2	-1	-1	-2	-4	-6	-7	-8	-9
DY	-12	-10	-7	-5	-4	-4	-4	-3	-3	-3
$W\gamma$	-3	-3	-1	-0	1	1	1	1	2	3
Wjet	-8	-9	-6	-5	-3	-0	-2	-1	-2	-1

8.5 Summary

In the search for the Standard Model (SM) Higgs in 1.9fb^{-1} of CDF data, there are 522 candidates observed with an expectation of 507±41 background events and 7.8±0.6 signal events for Higgs mass 160GeV/c^2 at next-to-next-to-leading logarithmic level calculation [64]. There are no significant excess found in the region sensitive to Higgs. The observed 95% C.L. upper limit at m_H = 160 GeV/c^2 is 0.8 pb which is 2.0 times the SM prediction while the median expected limit is $3.1^{+1.3}_{-0.9}$ with systematics included. Results for 9 other Higgs mass hypotheses ranging from 110GeV/c^2 to 200GeV/c^2 are also presented. The improved sensitivity by using Matrix Element method is equivalent to an effective luminosity increase by a factor of 3.4.

Chapter 9

Strong Evidence of the Standard Model ZZ Production

This chapter describes the results of ZZ search with 1.9 fb^{-1} data. The same selections and Matrix Element technique as in Higgs analysis are used. Section 9.1 reviews the event selections and shows the expected and observed yields. Section 9.2 defines the likelihood ratio (LR) variable calculated from WW and ZZ event probabilities. The frequentist approach is adopted in the ZZ analysis to calculate p-value and the equivalent Gaussian significance. The sensitivity of LR is also compared to other kinematic variable. Section 9.3 shows the special concern of using Vegas Monte Carlo integration instead of Importance Sampling to reduce the uncertainty of calculating event probabilities. Section 9.4 reports the observed significance and the measured cross-section. The results are combined with $ZZ \to 4l$ channel and submitted to PRL at the time we complete this thesis [8]. Section 9.5 tries to answer some statistical questions relating to our final results. Section 9.6 gives a brief summary.

9.1 Observed Yields and Systematics

ZZ candidates are constructed from events with exactly two identified leptons and opposite-sign. The event selections are identical to dilepton selection for $H \to WW$ analysis with two additional cuts: one is passing the cut $\not\!\!E_T^{\prime\,\mathrm{sig}} > 2.5\sqrt{\mathrm{GeV}}$ and the other is requiring same-flavor where CrkTrk is counted as either flavor. The phase space that satisfies these set of cuts is called ZZ region. The resulting background and signal expectations and data observations for the ZZ region are shown in Table 9.1.

The expected signal and background yields and observed yields are 256 ± 21(stat) and 276. More than half of the contribution is WW and the expected signal ZZ yield is 13.5 ± 1.1(stat).

There are several kinematics plots shown in Figure 9.1. There we have computed two shape hypothesis tests: Confidence Level test and Kolmogorov-Smirnov test (KS test) [79]. Figure 9.1(d) shows the invariant mass of two leptons. ZZ events are mostly distributed in the Z peak. The relevant backgrounds in this phase-space are WW, WZ and Drell-Yan. The size of each contribution is expected to be similar. There is an observation of deficit of data in the Z peak. The statistical interpretation will be discussed further in Section 9.5.

The shaded regions in Figure 9.1 are systematics uncertainties summarized in Table 9.2. The systematics are derived following the same procedure described in Chapter 7. The dominant and second dominant systematics are originated from the uncertainties of theoretical cross-section and NLO acceptance.

Table 9.1: Expected and observed yields in the ZZ region.

ZZ Fit Region

Category	WW	WZ	ZZ	$t\bar{t}$	DY	$W\gamma$	W+jets	Total	Data
e e	43.7	4.8	5.4	2.7	8.7	24.8	19.3	109 $\pm$10	118
e μ	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0 $\pm$ 0	0
μ μ	33.7	3.7	4.4	2.4	7.0	0.0	2.7	54 $\pm$ 5	45
e trk	35.3	2.3	2.2	2.4	3.8	5.9	9.9	62 $\pm$ 5	69
μ trk	19.2	1.5	1.5	1.4	1.5	1.1	5.2	31 $\pm$ 3	44
Total	131.8	12.3	13.5	9.0	21.1	31.7	37.1	256 $\pm$21	276

Table 9.2: Summary of the systematics uncertainties for ZZ analysis.

	Fractional Uncertainty (%)						
	WW	WZ	ZZ	$t\bar{t}$	DY	$W\gamma$	W+jets
$\not\!\!E_T$ Modeling	1.0	1.0	1.0	1.0	20.0	1.0	-
Conversions	-	-	-	-	-	20.0	-
NLO Acceptance	6.9	10.0	10.0	10.0	5.0	10.0	-
Cross-section	10.0	10.0	10.0	15.0	5.0	10.0	-
PDF Uncertainty	1.9	2.7	2.7	2.1	4.1	2.2	-
LepId $\pm 1\sigma$	1.4	1.4	1.4	1.4	1.4	1.0	-
Trigger Eff	2.5	2.2	2.1	2.5	3.4	6.1	-
Total	12.6	14.7	14.7	18.4	21.9	25.4	19.8

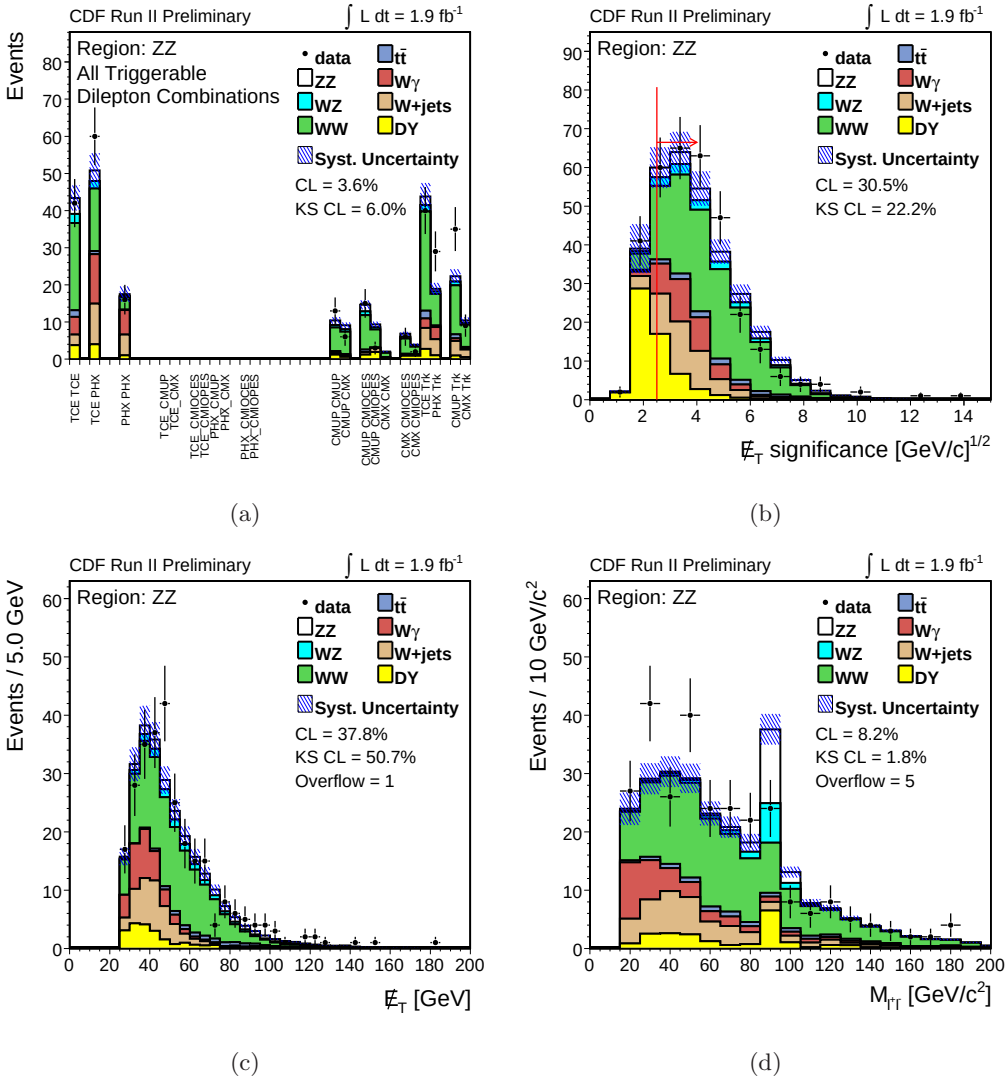


Figure 9.1: Signal and background distributions in the ZZ region.

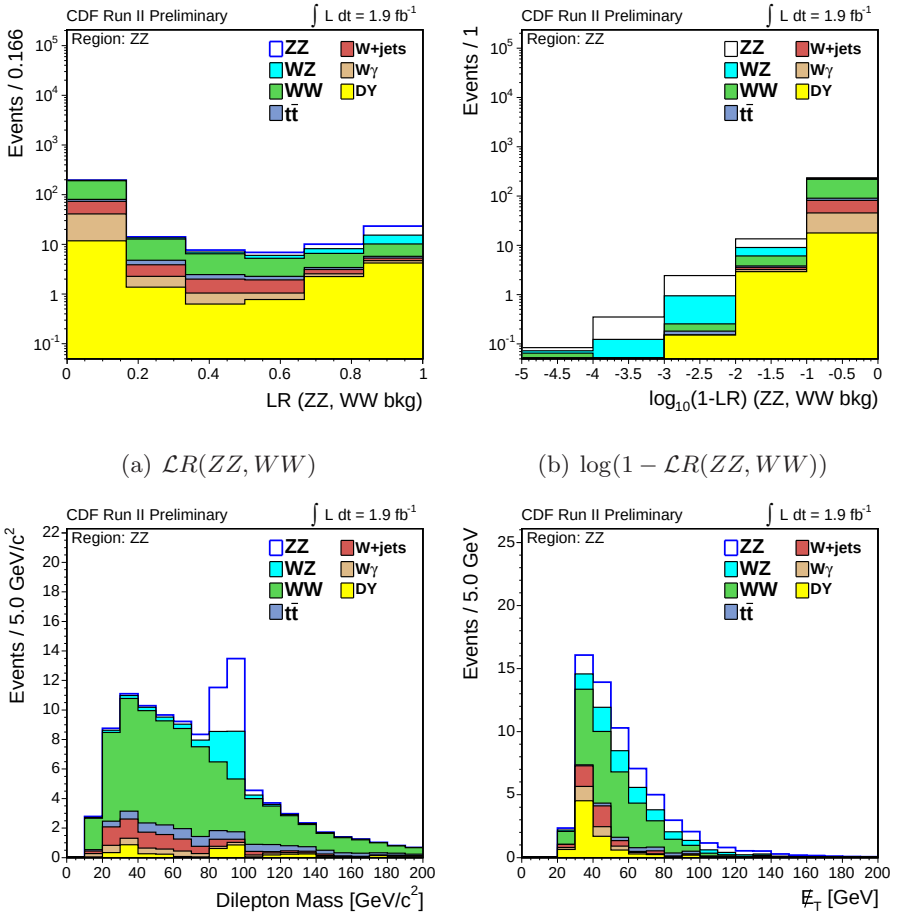
9.2 Likelihood Ratio Observable and Sensitivity

In this analysis, we apply the same idea as in $H \rightarrow WW$ analysis to use event probabilities for discriminating ZZ and WW. We construct the likelihood ratio:

$$\mathcal{LR}(ZZ, WW) = \frac{P_{ZZ}}{P_{WW} + P_{ZZ}}, \tag{9.1}$$

where P_{ZZ} and P_{WW} are event probabilities of ZZ and WW, respectively. We only use ZZ and WW event probabilities since $W\gamma$ and W+jets are irrelevant in ZZ-rich region as shown in Figure 9.1(d). The potential improvement of this likelihood ratio variable

is to take WZ event probability into account. The $\mathcal{L}R(ZZ, WW)$ distribution is shown in Figure 9.2(a). Because the likelihood ratio is so peaked near 1 for ZZ, we use the variable $\log(1 - \mathcal{L}R(ZZ, WW))$ to avoid binning away information. Figure 9.2(b) shows the distribution of this new variable. The other two variables which are sensitive to discriminate ZZ from WW are also shown in Figure 9.2(c) and Figure 9.2(d). They are invariant mass of dilepton with requiring $\not\!\!E_T > 50$ GeV and $\not\!\!E_T$ with requiring $M_{l^+l^-}$ within $76 \sim 106$ GeV/c^2, respectively.



(a) $\mathcal{L}R(ZZ, WW)$

(b) $\log(1 - \mathcal{L}R(ZZ, WW))$

(c) $M_{l^+l^-}$ in $\not\!\!E_T > 50$ GeV

(d) $\not\!\!E_T$ in Z mass

Figure 9.2: Kinematic obsesrvables for ZZ search.

9.2.1 Frequentist Approach

We use frequentist approach to calculate the p-value of observation and report the equivalent Gaussian significance. The p-value is defined as the probability of obtaining the test statistic at least as extreme as the one that was actually observed, given that the null hypothesis is true. The calculation is done by using package [73] and the procedures are described in this section.

The procedure this follows is to perform two likelihood fits: a null-hypothesis with the ZZ contribution fixed to zero and a test-hypothesis with the ZZ contribution floating. The test statistic is defined as

$$ts = (-2\ln\mathcal{L}_{ZZ\ free}) - (-2\ln\mathcal{L}_{ZZ\ fixed}) \tag{9.2}$$

where the likihood used is the same as Equation (6.13). The yield of each background in Equation (6.13) is allowed to float within systematic uncertainties in the fit. The observable is one of the kinematic variables in Figure 9.2.

We generate a large number (100K) of pseudo-experiments with background only (null hypothesis) assumption and calculate a test statistic from each of them. The p-value is defined as the probability of pseudo-experiments with ts smaller than observed ts_{obs} (more signal like):

$$\text{pvalue} = \text{Prob}(ts < ts_{obs}) = \frac{N(ts < ts_{obs})}{N_{p.e.}}. \tag{9.3}$$

The equivalent one-sided Gaussian significance is

$$\text{N}\sigma = 2 \times \text{erf}^{-1}(2 \times \text{pvalue}), \tag{9.4}$$

where error function erf is defined as

$$\text{erf(x)} = \frac{2}{\sqrt{\pi}} \int_0^{\text{x}} \text{e}^{-\text{t}^2}\text{dt}. \tag{9.5}$$

The significance is a convention which gives an intuitive picture to describe the area beyond N σ in one side of Gaussian distribution. The black solid histogram in Figure 9.3 shows test statistic distribution of the observable, $\log(1 - \mathcal{L}R(ZZ, WW))$, for pseudo-experiments with background assumption. The arrows indicates the values of test statistics to be 1σ, 2σ or 3σ which are -0.9, -3.8 and -8.6, respectively.

The red histogram in Figure 9.3 shows the test statistic distribution of the observable, $\log(1 - \mathcal{L}R(ZZ, WW))$, for pseudo-experiments with the SM (background

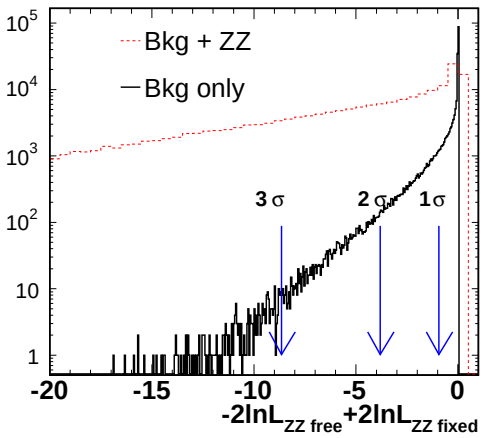


Figure 9.3: Test statistic distributions of pseudo-experiments with null and test hypothesis assumptions for the observable: $\log(1-\mathcal{L}R(ZZ,WW))$.

and ZZ) assumption. It can be used to calculate the probability of observing at least Nσ, e.g. the probability of observing more than 2σ is the fraction of entries less than -3.8 in the red histogram. Table 9.3 lists the probabilities of observing at least 2σ , 3σ or

Table 9.3: The probabilities of observing $ZZ \to ll\not{E}_T$ with at least 2, 3 or 5σ.

	$\log(1-\mathcal{L}R(ZZ,WW))$	$\mathcal{L}R(ZZ,WW)$	$M_{ll}(\not{E}_T > 50$ GeV)	$\not{E}_T(Z$window)
2 σ	0.54	0.57	0.49	0.49
3 σ	0.26	0.32	0.25	0.27
5 σ	0.06	0.06	0.03	0.03

5σ for using different observables based on 100k pseudo-experiments. It shows that we have a half-half chance to observe ZZ with at least 2 σ significance for each observable. However, the probability to observe at least 5 σ is doubled by using $\mathcal{L}R(ZZ,WW)$ related variables. The discrimination powers between $\log(1-\mathcal{L}R(ZZ,WW))$ and $\mathcal{L}R(ZZ,WW)$ are equivalent. We decide to use $\log(1-\mathcal{L}R(ZZ,WW))$ to report final results. It has an advantage to sort most ZZ-like events in order, the smaller the score it is, the more ZZ-like it is.

9.3 Vegas Integration

In this analysis, the observable used to construct the test statistic is log likelihood ratio:

$$\log(1 - \mathcal{L}R(ZZ, WW)) = \log(1 - \frac{P_{ZZ}}{P_{WW} + P_{ZZ}}), \tag{9.6}$$

where P_{ZZ} and P_{WW} are the calculated probabilities of ZZ and WW, respectively. To use this variable, it requires the accuracy of $\mathcal{L}R(ZZ, WW)$ calculations to $O(10^{-5})$. The Importance Sampling Monte Carlo integration method is not accurate enough. The score of $\log(1 - \mathcal{L}R(ZZ, WW))$ is sensitive to the seeding of the Monte Carlo integration and is quantified in this section. We use Vegas integration algorithm as alternative way to calculate log likelihood ratio. The improvement of the accuracy of the log likelihood ratio is quite significant. Vegas algorithm optimizes the integration phase-space such that the variations of weights for each Monte Carlo integration points are relatively equal. It reduces the statistical uncertainties which has $\sqrt{\sum w^2/N}$ dependences. To have an idea of computing time, the 40k importance sampling takes 4 seconds per event but it takes 68 seconds for Vegas Integration. The test is done with CPU Pentium Duo 2 GHz and 2 GB RAM. The final integration output is to use 100k Vegas sampling points.

Figure 9.4 shows the distributions for the three most ZZ-like candidate events with different MC integration configurations. The red and blue arrows are calculation outputs by using Importance Sampling with two different seeds. The pink histogram shows the distribution of log-likelihood ratio values due to varying the seed of the importance sampling method with 40k sampling points for each integration. The size of variation could be as large as 1.5 bin size. It reflects large uncertainty in the WW event probability because the importance sample rarely populates the corners of the WW phase space that produce ZZ like events. The green arrows are results to report final results by using Vegas integration. The black histogram is the variation due to different seedings with Vegas algorithm. The size of variation is less than a quarter of bin size.

9.4 Results

The p-values listed below are based on 10 million toy experiments counting how many experiments have a smaller (more signal like) test statistic than the one observed in data. The resulting $\log(1 - \mathcal{L}R(ZZ, WW))$ distributions along with other sensitive variables are shown in Figure 9.5. The results are combined with four lepton

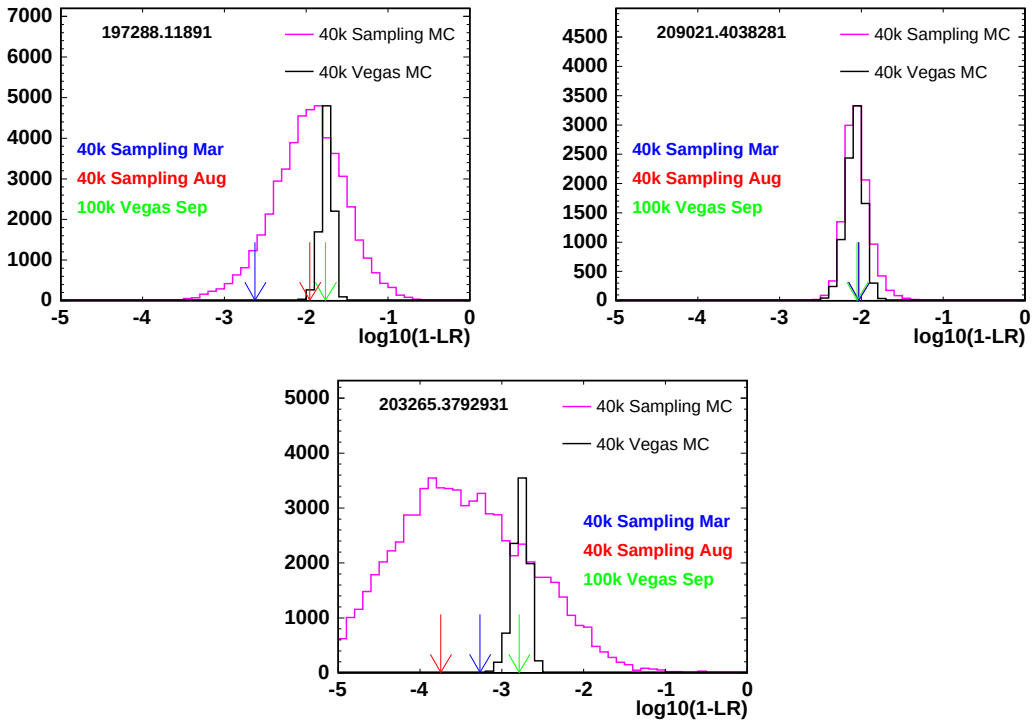
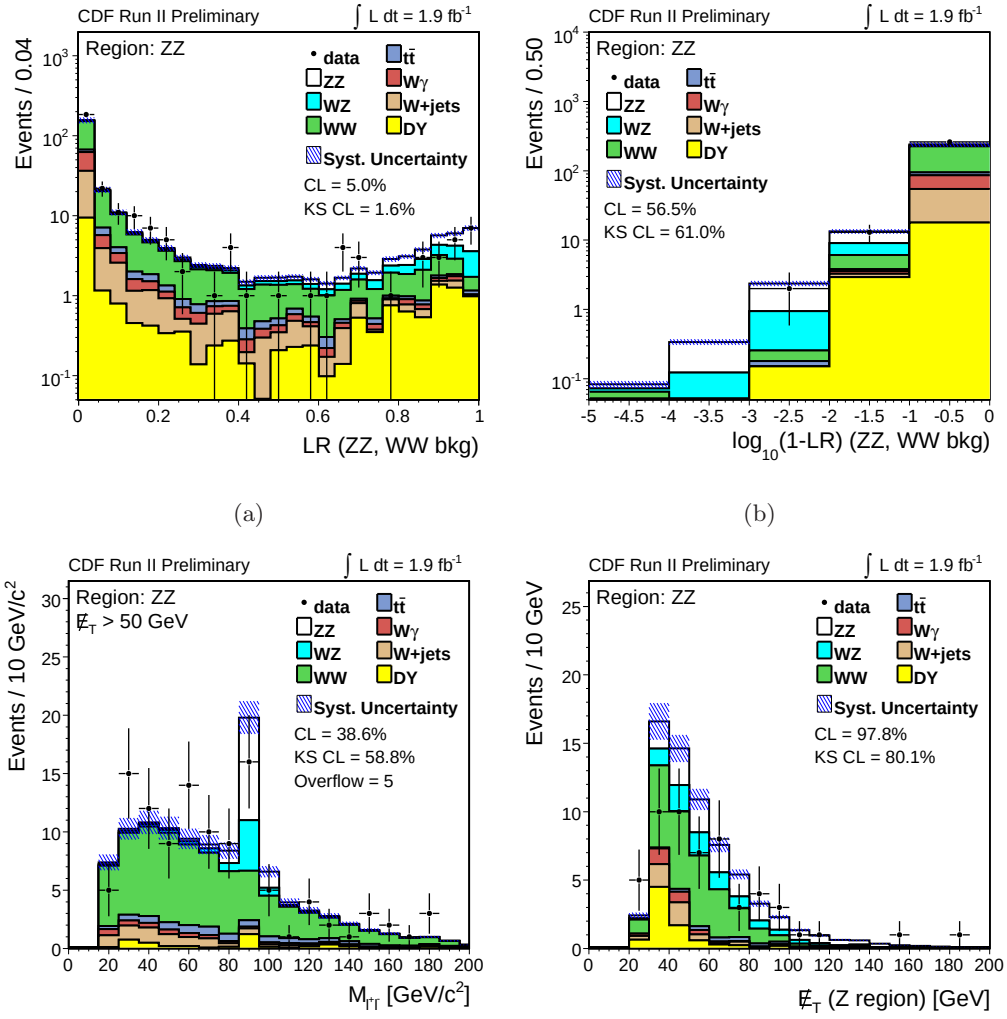


Figure 9.4: The $\log(1-\mathcal{L}R(ZZ,WW))$ value for the three most ZZ-like candidate events. The final result is the green arrow which use 100k Vegas Integration.

Table 9.4: The p-value and significance of observing $ZZ \to ll\not{E}_T$ and $ZZ \to 4l$.

Discovery Probability	$ll\nu\nu$	4 lepton	combined
2σ	0.55	0.82	0.87
3σ	0.33	0.67	0.75
5σ	0.06	0.34	0.50
p-value	0.1177 ± 0.0001	$(1.09 \pm 0.10) \times 10^{-5}$	$(5.10 \pm 0.07) \times 10^{-6}$
observed significance	1.2 σ	4.2 σ	4.4 σ

channel described in [8] and shown in Table 9.4. The same systematics (e.g. cross-section, luminosity, etc.) are treated as correlated between the two lepton and four lepton channels. We have a 55% chance to see an excess in the $ll\not{E}_T$ channel that has a statistical significance of more than 2σ, assuming a cross section of 1.4 pb. The data observation in $ll\not{E}_T$ channel is 1.2σ while the total significance including $4l$ is 4.4σ.



(a)

(b)

(c) m_{ll} for $\not\!\!E_T > 50$ GeV ($\approx$ the most sensitive region)

(d) $\not\!\!E_T$ for $76 < m_{ll} < 106$ GeV

Figure 9.5: Kinematics distributions of ZZ candidates. The shaded histogram shows systematics uncertainty.

Cross-section Determination

The cross-section is determined from the same fit. Because the systematics are included as Gaussian constraints (including the correlations between systematics in different channels) the fit uncertainty includes both the statistical and systematic uncertainties. The results measured cross-section is

$$1.37^{+0.74}_{-0.58}(stat. + syst.) \text{ pb} \tag{9.7}$$

which should be compared to the NLO prediction in the zero-width approximation of 1.4 pb [65].

9.5 Discussion

In this section, we are going to answer two statistical questions in $ZZ \to ll\not\!\!E_T$ analysis. We will focus on the ZZ-rich region, $76 < m_{ll} < 106$ GeV/c^2. The first question is to ask what's the probability of the yield difference greater than what we observed in the data. The second question is what's the probability of seeing the observed significance while we have deficit as we seen in the data.

Table 9.5 shows the expected yields in $76 < m_{ll} < 106$ GeV/c^2 is 68 ± 9 while we observed 54. We do 1M toy experiments with given predicted yields and consider the correlations between different systematics and calculate the probability is 49%. Fig 9.6 shows the results of pseudo-experiments. Table 9.6 shows the break down of systematics due to different sources. The major systematics are coming from NLO acceptance and theoretical uncertainties. Consider the good consistency outside Z mass window, it's quite unlikely that the large yield fluctuation in this small Z mass window are caused by these two sources. We then evaluate the probability by ignoring these two systematics and the result is 39% as shown in the right plot of Fig 9.6. The major systematics is $\not\!\!E_T$ which is basically the fake $\not\!\!E_T$ model of Drell-Yan.

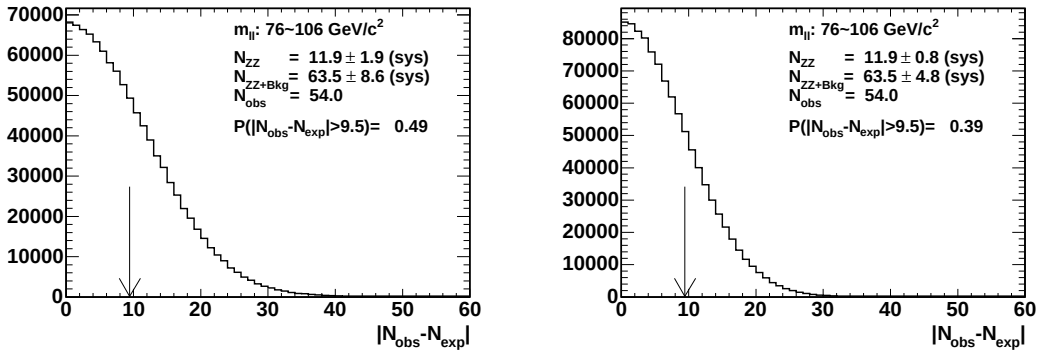
Table 9.5: Summary of the yield systematics in the Z window.

Process	Yields	Systematics	Fractional (%)
WW	25.30	3.4	13.3
WZ	8.90	1.4	15.8
ZZ	11.92	1.9	15.8
$t\bar{t}$	1.90	0.4	19.3
DY	8.30	1.9	22.7
$W\gamma$	2.60	0.7	26.3
W+jets	4.60	1.0	22.5
Total	63.5	8.6	13.5
Observed	54		

The other question we could ask is what's the probability of seeing the observed significance while we have deficit as seen in the data. The probability is calculated to be 44% by doing pseudo-experiments. The procedure is as follow. For each pseudo-experiment, we determined the number of events for each process by Poisson the expected

Table 9.6: Break down of the different sources of the yield systematics in the Z window.

Sources	Systematics	Fractional (%)
NLO Acceptance	4.3	6.8
Cross-section	5.6	8.8
$\not\!\!E_T$ Modeling	2.2	3.4
Conversions	0.5	0.8
PDF Uncertainty	1.5	2.3
LepId $\pm 1\sigma$	0.8	1.3
Trigger Eff	1.5	2.3
Luminosity	3.5	5.6
Fake	1.0	1.6
Total	8.6	13.5



(a) (b)

Figure 9.6: The probability of the yield difference greater than what we observed is 49% with full systematics (left) 9.6(a) and 39% while ignoring NLO and cross-section systematics 9.6(b).

yields with given systematics in the ZZ signal region. We get un-weighted events by going through each process after we determined the yields. The next step is to check the number of events in the Z mass region $76 < m_{ll} < 106$ GeV/c^2. We only keep the pseudo-experiments whose Z mass region yields are less than the observed yields of 54. The same statistical fitter is applied on these experiments to evaluate the significance. Figure 9.7 shows the distributions of significance from these pseudo-experiments.

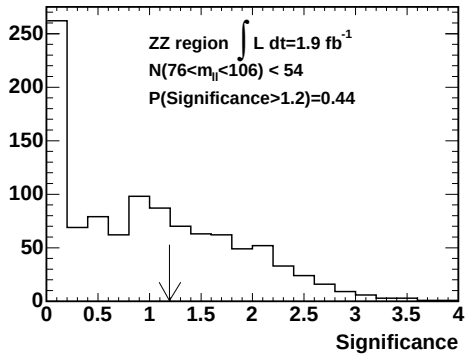


Figure 9.7: The probability of significance greater than 1.2 while we have yield deficit less than 54 is 44%.

9.6 Summary

We have described a search for ZZ production using two lepton plus two neutrino final state. The result is combined with the four lepton channel in [8]. We observed an excess yield that has a statistical significance of 4.4 σ after systematics are taken into account. The corresponding measured cross-section is $1.37^{+0.74}_{-0.58}(stat. + syst.)$ pb which should be compared with a Standard Model expected cross-section at NLO of 1.4 pb when requiring both Z are on-shell.

Chapter 10

WW Cross-Section Measurements

This chapter describes the results of WW cross-section measurement with 1.9fb^{-1} data. The same lepton selections and event selection as Higgs analysis has been used. A cross-section measurement is performed and the possible improvement of using event probability to constrain triple gauge coupling parameters are shown. Section 10.1 describes the measurement of WW cross-section. Section 10.2 describes the determination of triple gauge coupling limits. Section 10.3 gives a brief summary of the results.

10.1 Cross-Section Measurement

Acceptance

The Monte Carlo sample used to estimate the WW acceptance is described in Section 5.1. The baseline estimate of the efficiency is made separately for each of the lepton pair category described in chapter 4. Each category is corrected for the lepton identification scale factors and trigger efficiencies that are relevant to that specific pair and for the relative luminosity of applicable good run list to the em_nosi good run list. The master formula is then

$$A_i = \frac{N_i^{\rm reco}}{N_i^{\rm gen}(|Z_0| < 60 \text{ cm})} \times \epsilon_i^{trg} \times s_i^{lep} \times \epsilon_{vtx} \times \frac{\mathcal{L}_i}{\mathcal{L}_{\rm em\_nosi}} \quad (10.1)$$

where A_i is the acceptance times the efficiency and

- $N_{\rm reco}$ is the number of reconstructed WW events passing the full selection,

- $N_{\rm gen}(|Z_0| < 60$ cm) is the number of generated events within 60 cm of the interaction point
- ϵ_{trg} is the effective trigger efficiency for the category (see below),
- s_{lep} is the effective lepton id scale factor (see below),
- ϵ_{vtx} is $0.9555 \pm 0.0004(stat) \pm 0.0031(sys)$ from CDF note 7935, and
- $\frac{\mathcal{L}_i}{\mathcal{L}_{\rm em\_nosi}}$ accounts for the relative amount of the applicable good run list.

Once calculated, the summary acceptances for the ee, $e\mu$, and $\mu\mu$ categories are shown in Table 10.1. The acceptances defined with the luminosity ratios are truly comparable in terms of predicting expected events and converting an observed number of events to a cross-section. The number is normalized to the total WW rate, so it includes the W branching fractions, possible τ branching fractions, and detector acceptance and reconstruction efficiency. This acceptance is 50% larger than the previous Run 2 WW analysis [5]. There are 30% coming from adding new leptons and another 21% coming from including 1 jet.

Table 10.1: The expected yields and acceptance for WW after Base selection.

Cut	ee	$e\mu$	$\mu\mu$	et	$\mu\ t$	$ll(\pm stat)$
TrigDil	81.0	154.9	61.4	65.5	35.6	398.4 ± 1.0
maxΔz	81.0	154.8	61.4	65.4	35.6	398.2 ± 1.0
m_{ll}	80.0	152.7	60.3	64.6	35.1	392.8 ± 1.0
$\not\!\!E_T^{spec}$	51.4	118.4	37.5	40.6	21.5	269.5 ± 0.9
OS	48.4	114.8	37.5	39.3	21.5	261.6 ± 0.8
Njets<2	46.6	110.1	36.0	37.8	20.6	251.0 ± 0.8
0jet	37.0	87.5	28.6	30.1	16.3	199.5 ± 0.7
A(Njets<2)	2.0e-03	4.7e-03	1.5e-03	1.6e-03	8.7e-04	$1.065e-02 \pm 3e-05$

Results

Table 10.2 lists the observed and expected yields for WW and background. We find 522 events expecting a signal of $N_{\rm sig} = 251.0 \pm 0.8 \pm 30$ events, with a background of $N_{\rm Bkg} = 256.2 \pm 2.6 \pm 28$. The signal expectation corresponds to an acceptance of $A = (101.2 \pm 0.1 \pm 4.2) \times 10^{-4}$ where this acceptance includes all efficiency factors and branching fractions. The equation used to extract the final cross-section is

$$\sigma = \frac{N_{\rm Data} - N_{\rm Bkg}}{A\ \mathcal{L}}, \tag{10.2}$$

Table 10.2: Expected yields of WW, total backgrounds, W+jets and observed data. The luminosity uncertainty is not included in the systematics.

Category	WW	Background	W+jets	Data
ee	$46.6 \pm 0.1 \pm 6$	$90.6 \pm 1.6 \pm 10$	$22.8 \pm 1.0 \pm 5$	144
$e\mu$	$110.1 \pm 0.3 \pm 13$	$80.9 \pm 1.0 \pm 10$	$24.1 \pm 0.9 \pm 5$	191
$\mu\mu$	$36.0 \pm 0.1 \pm 4$	$32.3 \pm 0.8 \pm 4$	$3.2 \pm 0.1 \pm 1$	58
et	$37.8 \pm 0.1 \pm 4$	$35.3 \pm 0.8 \pm 4$	$11.0 \pm 0.6 \pm 2$	80
μt	$20.6 \pm 0.0 \pm 2$	$17.1 \pm 0.4 \pm 2$	$5.8 \pm 0.3 \pm 1$	49
Total	$251.0 \pm 0.8 \pm 30$	$256.2 \pm 2.6 \pm 28$	$66.8 \pm 1.7 \pm 15$	522

where $N_{\rm Data}$ is the observed yield in data and $\mathcal{L}$ is the luminosity. The luminosity uncertainty is applied to both the denominator and the luminosity dependant background, $N_{\rm BkgLumDep} = 189.3$ (all except fakes). The formula for this is

$$(\Delta\sigma)_{\mathcal{L}} = \Delta(\frac{N_{\rm Data} - N_{W+\rm jets} - N_{\rm BkgLumDep}}{A\ \mathcal{L}}) = \frac{N_{\rm Data} - N_{W+\rm jets}}{AL}\frac{\Delta\mathcal{L}}{\mathcal{L}} \tag{10.3}$$

This yields the final cross-section

$$\sigma(p\overline{p} \rightarrow WW) = 13 \pm 1(stat) \pm 2(sys) \pm 1(lum)\ pb \tag{10.4}$$

10.2 Anomalous Triple-Gauge-Couplings

The study in this section is to explore the possible improvements by using event probability. There are multiple coupling schemes for WW aTGC. The HISZ scheme[31] is used as demonstration purpose. A comparison of constraint sensitivity with a one-dimensional observable is done. The considered variable, M_T, is driven from the production cross-section enhancement at high $\sqrt{s}$. Due to missing neutrinos, the total center of mass energy is not possible to reconstruct but M_T is highly correlated to $\sqrt{s}$.

The parametrization of the charged triple gauge boson vertices is described in References. The most general Lorentz invariant Lagrangian which describes the triple gauge boson interaction has fourteen independent complex couplings, seven describing the $WW\gamma$ vertex and seven describing the WWZ vertex. Assuming electromagnetic gauge invariance as well as C and P conservation, the number of independent TGCs reduces to five. A common set is $g_1^Z, \kappa_Z, \kappa_\gamma, \lambda_Z, \lambda_\gamma$ where $g_1^Z = \kappa_Z = \kappa_\gamma = 1$ and $\lambda_Z = \lambda_\gamma = 0$ in the Standard Model. The parameters proposed in [31] and used by the LEP experiments are g_1^Z, λ_γ and κ_γ with the gauge constraints:

$$\kappa_Z = g_1^Z - (\kappa_\gamma - 1)tan^2\theta_W \tag{10.5}$$

$$\lambda_Z = \lambda_\gamma, \tag{10.6}$$

where θ_W is the weak mixing angle. The couplings are considered as real, with the imaginary parts fixed to zero. Note that the cTGC parameters can be quoted in different notations. We can fit three parameters $\Delta g_1^Z = g_1^Z - 1$, $\Delta\kappa_\gamma = \kappa_\gamma - 1$, and $\lambda_\gamma = \lambda_Z$ and connect the other two variables through relation, $\Delta\kappa_Z = \kappa_Z - 1 = \Delta g_1^Z - \Delta\kappa_\gamma tan^2\theta_W$.

Reweighting Technique

To prepare Monte Carlo sample at a given aTGC coupling value α, we can use re-weighting technique to save computing time. The idea is to assign a weight to each Monte Carlo event which are generated at SM point. The SM Monte Carlo is unweighted from leading order generator and passed through full detector simulation. The formula for reweighting is

$$N(\alpha) = \sum Nw = \sum_N \frac{f_i f_j |M_{ij}(\alpha)|^2}{f_i f_j |M_{ij}(SM)|^2}, \tag{10.7}$$

where M_{ij} is the matrix element and f_i is the parton density function. The validity of this equation can be observed by setting α to SM value such that the left hand side equals the SM prediction. This technique works quite well for WW aTGC Monte Carlo because the TGC parameter ranges of interest do not deviate too much from SM point. The phase spaces in larger $\sqrt{s}$ is not popular for SM such that the weighted uncertainty is larger in those regions.

Setting Confidence Interval

Figure 10.1 shows the expected yields at 1.9fb^{-1} as a function of aTGC parameter. The variation of each coupling is computed while setting the other coupling parameters to zero. Given the one σ statistical fluctuation of signal yields is $\sigma =\sim \sqrt{250} = 16$, the expected 95% C.L. limit is about 3 σ = 48. we can read the expected 95% C.L. limit by just using production cross-section information: $|\lambda_\gamma| < 0.25$, $|\delta g_1^Z| < 0.4$ and $|\delta\kappa_z| < 0.9$.

We construct likelihood function with the same definition as Higgs analysis. The Bayesian statistics is used to set 95% C.L. limit. The log likelihood function $-2logL(x)/L_0$ is shown in Figure 10.2(a) where L_0 is the maximum of log likelihood function. The log likelihood function value 1 typically means the 68% C.L. range and 2.95 means the 95% C.L. range. In the case of the minimum of log likelihood function is offset from zero, the log likelihood function will be distorted and non-parabolic

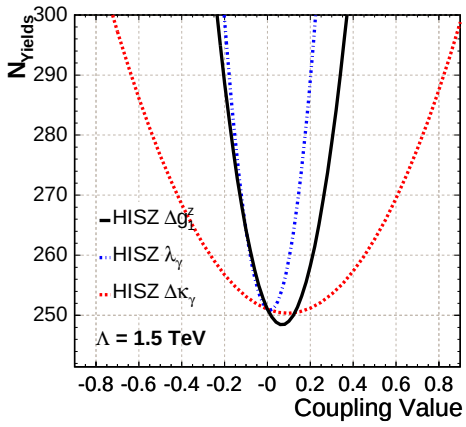


Figure 10.1: The expected yields at $1.9 fb^{-1}$ for different aTGC coupling values. The HISZ is used in this case.

shape, e.g. Figure 10.2(b). If the offset is larger than 1σ, the log likelihood function can have double minimum as shown in Figure 10.2(c) because the production cross-section of aTGC is close to quadratic. Instead of looking for the intersection of $-2logL/L_0 = 1$, the correct way to calculate the 95% C.L. interval is to integrate over likelihood function and find the intervals give us 95% probability as shown in Figure 10.2(d). We could have two island interval if the offset is large enough.

Sensitivity Comparison

Figure 10.3 shows the kinematics comparison of $\Delta\Phi(l^+l^-)$ and M_T at 1.9 fb^{-1}. Transverse mass M_T is defined [80] as

$$\begin{aligned} M_T^2 &= M_C^2 - p_T^2(l^+l^-\nu) \\ M_C &= \sqrt{p_T^2(l^+l^-) + m^2(l^+l^-) + p_T(\nu)}. \end{aligned} \tag{10.8}$$

Figure 10.4(b) shows the kinematics comparison of $\Delta\Phi(l^+l^-)$ and M_T at some selected anomalous coupling points. We can see the transverse mass distribution is enhanced at higher M_T region. This is explained by the introduction of aTGC via s-channel and larger $\sqrt{s}$ gives larger aTGC production rate. In high $\sqrt{s}$, two W is boosted further in the transverse plane and the decayed lepton tends to be boosted along the W moving direction such that two leptons tend to be back-to-back. This signature could be seen in the Figure 10.4(d) The two dimensional correlation of $\Delta\Phi(l^+l^-)$ and m_T is shown in the Figure 10.4(a) and Figure 10.4(c). The former is generated at SM and the later

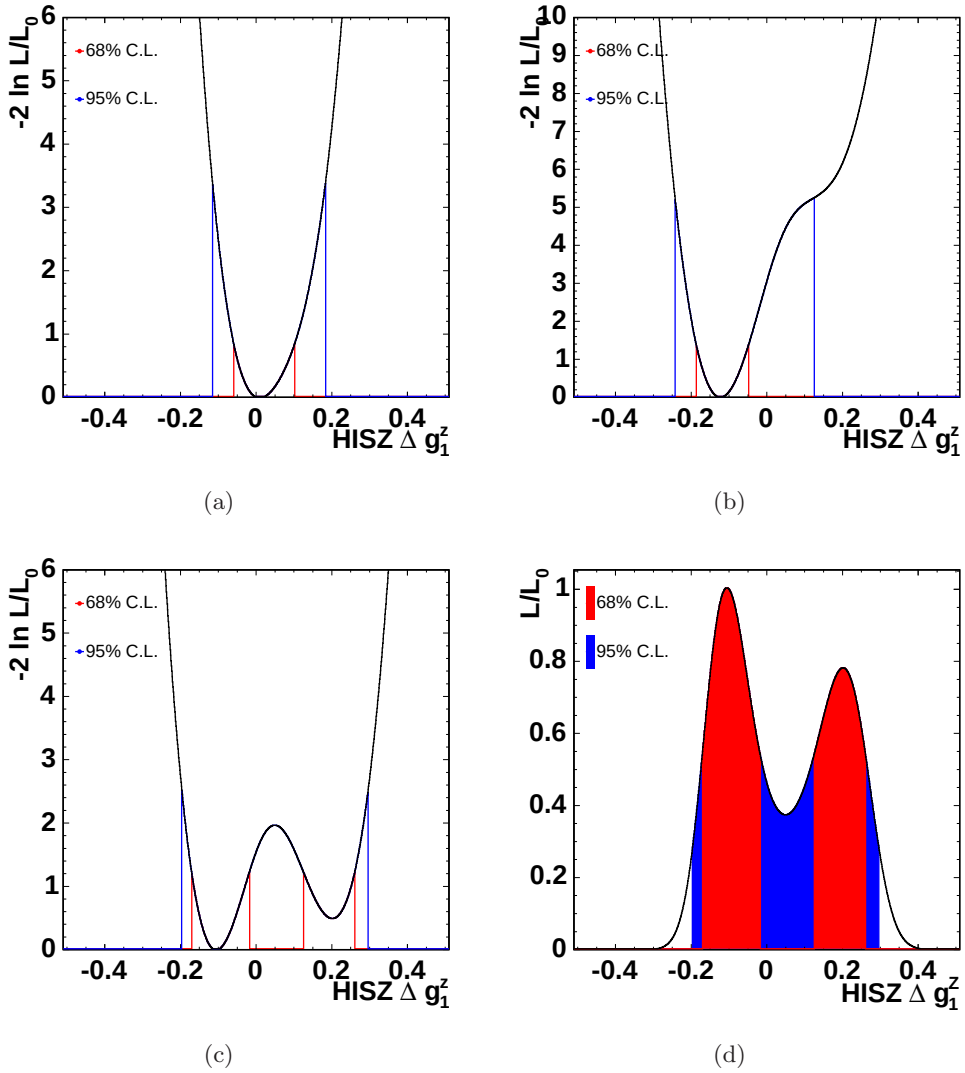
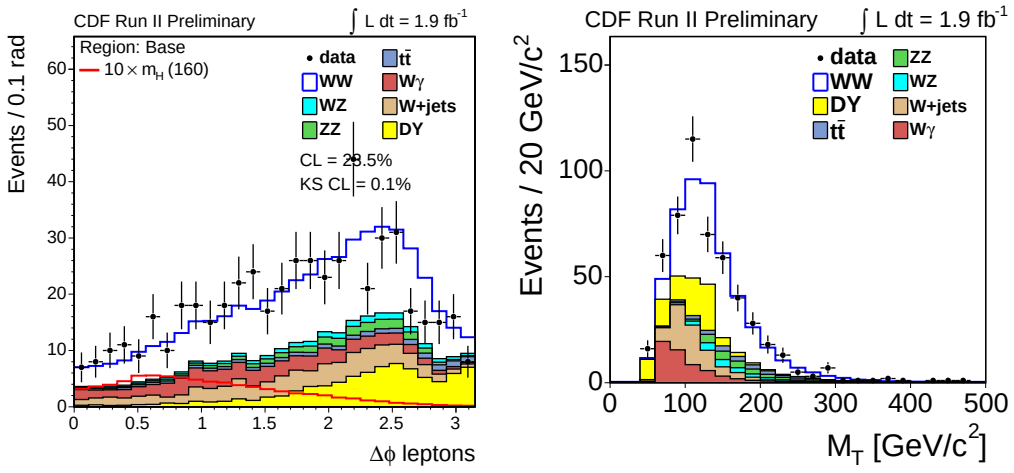


Figure 10.2: The typical likelihood functions show double local minimum.

is setting Δg_1^Z to 0.4. The comparison of 2D plots show more clear that the aTGC is populated in the high M_T and large $\Delta\Phi(l^+l^-)$ phase space.

To compare the sensitivity of using different observables, we generate 1000 pseudo-experiments assuming SM and calculate 95% C.L. intervals on each of the aTGC parameters. The pseudo-experiments are generated with MCFM and convoluted with detector resolution and efficiency. The acceptance is chosen to be $|\eta| < 2$ and $|\eta| < 1.5$ for electrons and muons. The $\not\!\!E_T^{spec}$ is required to be greater than 25 GeV and 15 $|gev$ for ee,$\mu\mu$ and $e\mu$ channel, respectively. The p_T requirements for two leptons are



(a) (b)

Figure 10.3: (a) $\phi(l^+l^-)$ (b) transverse mass distribution in 1.9 fb^{-1} in the Base region.

(20 GeV/c, 10 GeV/c). The backgrounds are not included in this study but does not affect our purpose to look for the best observables. We perform a maximum likelihood (ML) fit by using the event probability calculated via Matrix Element technique. This gives us the best sensitivity of aTGC limit. The template fitting is done on other observables and compared to the ML fit results.

The expected 95% C.L. intervals between different observables including ML fit are shown Figure 10.5. M_T alone gives the expected interval to be: $|\Delta g_1^Z| < 0.15$, $|\lambda_\gamma| < 0.09$ and $|\Delta\kappa_\gamma| < 0.37$. By including azimuthal angle for a two dimensional fitting M_T v.s. $\Delta\Phi(l^+l^-)$, the sensitivity is improved by 1% while it is 2% for ML fit. Consider the 3% statistical uncertainty of 1000 pseudo-experiment, we can conclude that the transverse mass fit has deliver the optimal sensitivity. The angular information does not buy us too much to separate aTGC from SM. The aTGC sensitivity depends mostly on center-of-mass energy $\hat{s}$ and it is smeared by parton density function.

10.3 Summary

We have measured production cross-section for WW with the same event selection as Higgs analysis and the result is $\sigma(p\overline{p} \to WW) = 13 \pm 1(stat) \pm 2(sys) \pm 1(lum)$ pb. The statistical uncertainty is smaller than the systematics. The aTGC coupling precision can reach 9% for λ_γ in HISZ scheme with energy cut-off $\Lambda = 1.5$ TeV. The observable

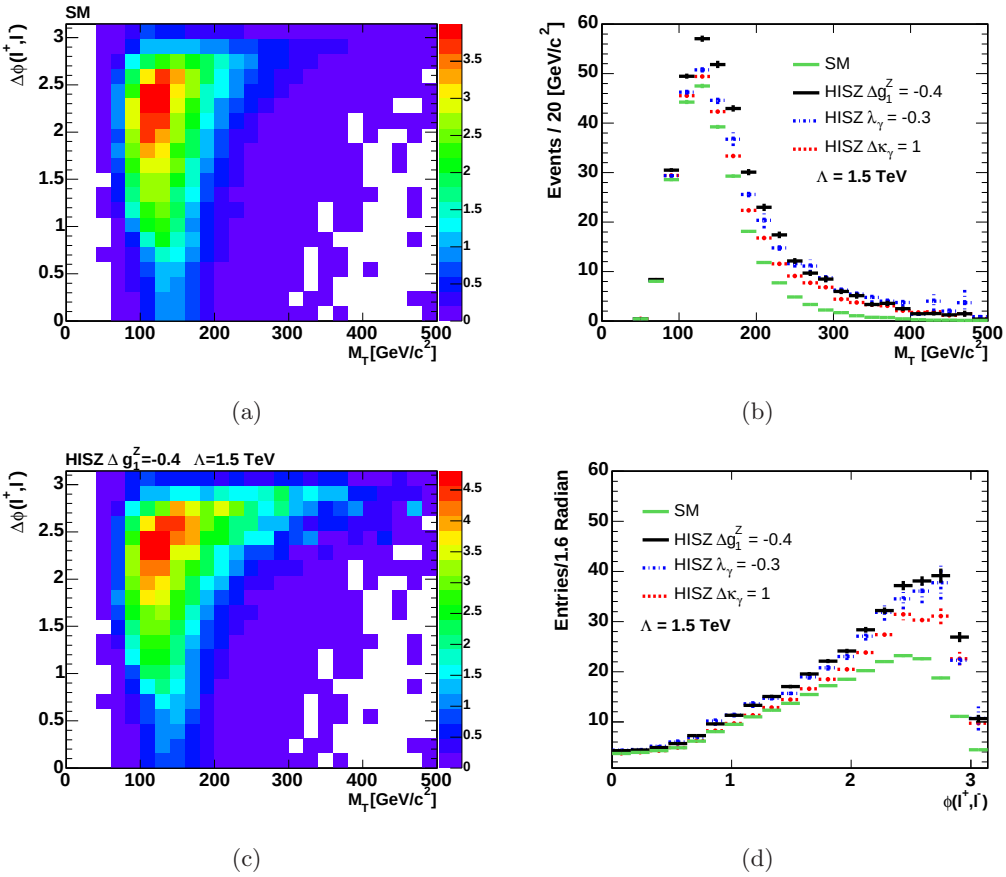


Figure 10.4: The transverse mass and azimuthal angle distribution at three different anomalous coupling points.

M_T has delivered as much as information as using event probability due to the fact that aTGC is driven by $\hat{s}$ and the production is averaged over parton PDFs.

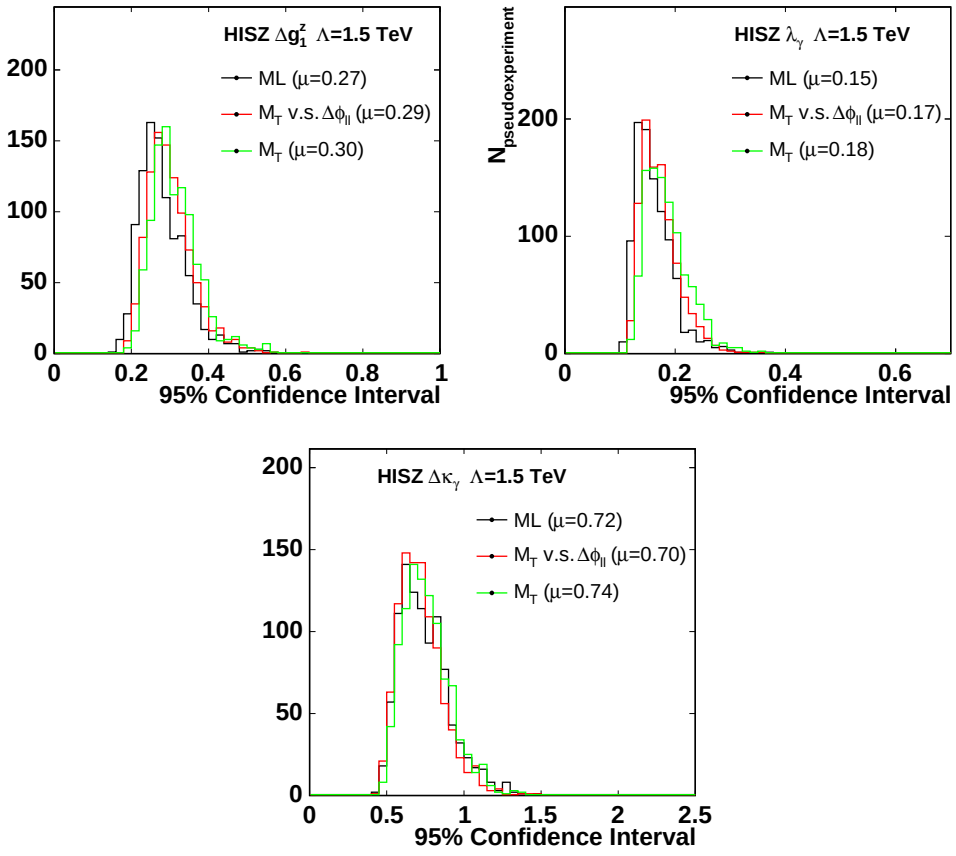


Figure 10.5: The expected 95% C.L. intervals for Δg_1^Z, $\delta\kappa_g$ and λ_g in HISZ scheme at energy cut off $\Lambda = 1.5$ TeV with respect to observables.

Chapter 11

Conclusion

We report on studies of the Standard Model production of Higgs, ZZ and WW in the two charged lepton (e, μ) and two neutrino final state. The data were collected with CDF II detector at the Fermilab Tevatron in $p\overline{p}$ collisions at a center of mass energy $\sqrt{s}$ = 1.96 TeV and correspond to 1.9 3 fb^{-1}. The improvement of the analysis are done in two aspects: for one the increase of acceptance by exploiting new lepton types and for two the improvement of sensitivity by developing Matrix Element techniques. The data used for Higgs search at the completion of this thesis is 1.9 fb^{-1} in June 2008. With additional 1 fb^{-1} data, the results used techniques described in this thesis is submitted to Physical Review Letter on Sep 23 2008 [10]. The measurement of ZZ production is published in Jan 2008 [8]. The new lepton types described in this thesis also contributed to WZ observation [9]. Figure 11.1 shows the current understanding of diboson production contributed from this thesis.

In the search of the Standard Model (SM) Higgs at 1.9 fb^{-1}, there are 522 candidates observed with an expectation of 507±41 background events and 7.8±0.6 signal events for Higgs mass 160GeV/c^2 at next-to-next-to-leading logarithmic level calculation. There is no significant excess found in the region sensitive to Higgs. The observed 95% C.L. upper limit is 0.8 pb at m_H = 160 GeV/c^2 which is 2.0 times the SM prediction while the median expected limit is $3.1^{+1.3}_{-0.9}$ with systematics included. Results for 9 other Higgs mass hypotheses ranging from 110GeV/c^2 to 200GeV/c^2 are also presented. The improved sensitivity by using Matrix Element method is equivalent to have effective luminosity increase by a factor of 3.4.

The measurement of ZZ production using a combination of the four lepton and two lepton plus two neutrino final states results in a 4.4 σ excess above backgrounds.

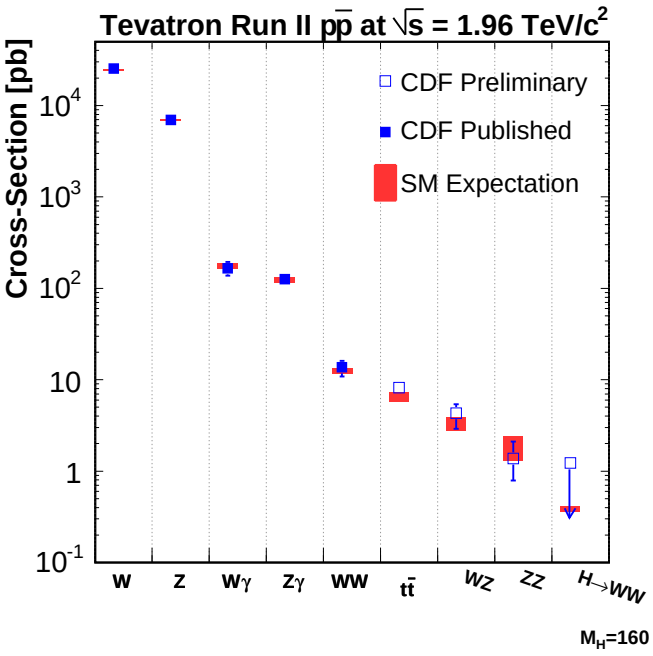


Figure 11.1: The experimental results of diboson production measurements at CDF in summer 2007. WW, ZZ and Higgs productions are contributed by this thesis.

The corresponding measured cross-section is $1.37^{+0.74}_{-0.58}(stat. + syst.)$ pb which should be compared with a standard model expected cross-section at NLO of 1.4 pb [65].

The measured production cross-section for WW is $\sigma(p\overline{p} \rightarrow WW) = 13 \pm 1(stat) \pm 2(sys) \pm 1(lum)$ pb which is consistent with the standard model expected cross-section at NLO of 12.4 ± 0.8pb [5]. The statistical uncertainty is smaller than the systematics. The aTGC coupling precision can reach 9% Confidence Interval for λ_γ in HISZ scheme with energy cut-off $\Lambda = 1.5$ TeV. The observable M_T has performed nearly the optimal observable in $p\bar{p}$ collision for WW triple-gauge-coupling study.

Appendix A

Trigger Efficiency

The basic idea for measuring the trigger efficiency is to measure the fraction of the number of events passing the trigger bit requirement of interest from a given signal event, e.g. a $Z \to l^+l^-$ or $W \to l\nu$ [81].

In the case of using $Z \to l^+l^-$, one of the lepton is passing through the tight lepton selection and paired with another lepton passing through the trigger level criteria. The lepton pair invariant mass is required to be in the Z-pole. The fraction of the probe leg lepton matched to the trigger object is measured efficiency. The trun-on curve of trigger efficiency for central electron and muon trigger path is very fast. The trigger efficiency is applied as a constant scale factor based on a run-by-run dependence as shown in Table A.1.

The MET_PEM trigger path requires large $\not\!\!E_T$ and also a large PEM E_T tower. Both W and Z are used to measure the trigger efficiency. The larger calorimeter resolution dilute the turn-on curve in raw data level such that the turn-on curve of trigger efficiency is slower as shown in the Figure A.1 The trigger efficiency is also measured with respect to offline $\not\!\!E_T$ and PEM E_T which is applied in the analysis level. Other

Table A.1: Averaged trigger effieciencies for period 0 (138425∼186598) and period 1-7 (190697∼212133).

	period 0	period 1-7
TCE	0.9620 ± 0.006	0.9766 ± 0.006
PHX	0.910 ± 0.003	0.910 ± 0.003
CMUP	0.8977 ± 0.005	0.9186 ± 0.004
CMX	0.9665 ± 0.004	0.9586 ± 0.004

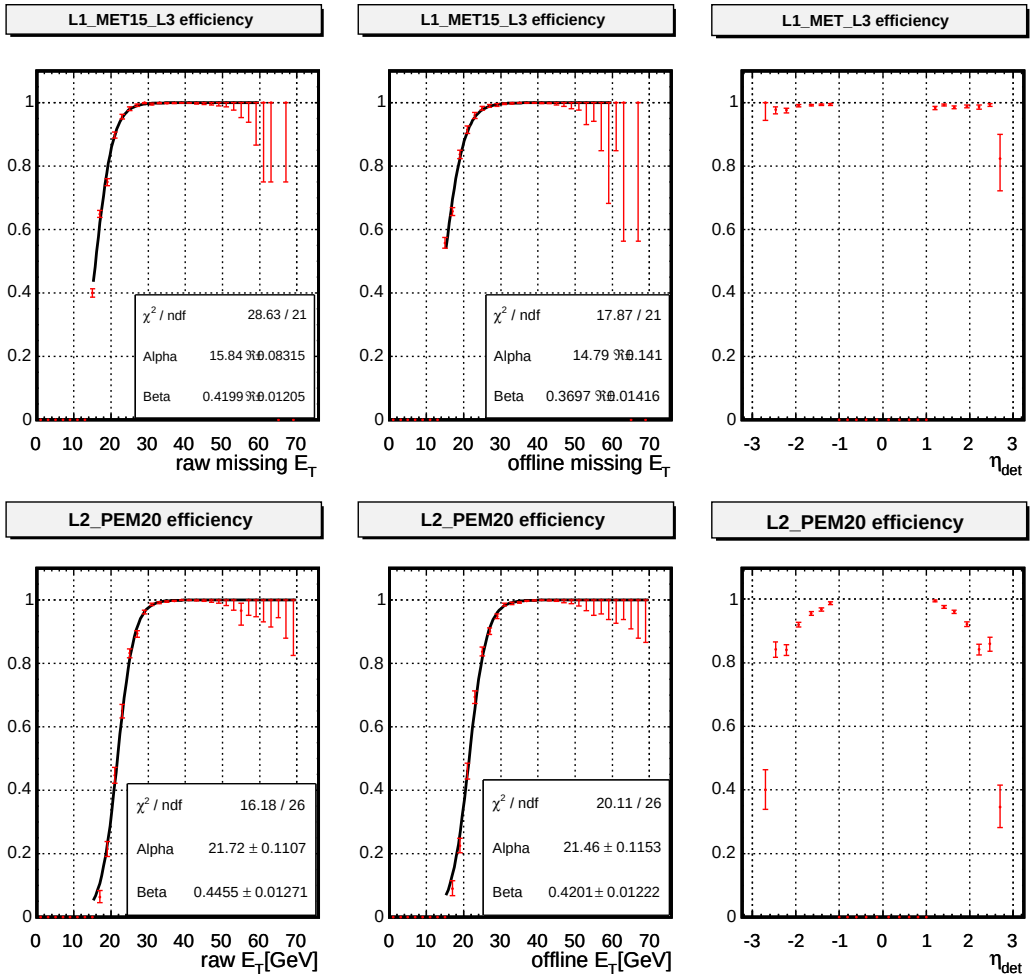


Figure A.1: The turn-on curve of MET_PEM trigger efficiency for Level 1 $\not\!\!E_T$ and Level 2 PEMET and the η dependence of L2_PEM20 trigger efficiency are all applied in this thesis [81].

than that, the η dependence efficiency of the L2PEM is also applied.

Appendix B

Lepton Identification

In order to measure the lepton efficiencies in data, we use Z candidates selected with one leg that has a full lepton selection (*tag*-leg) and one leg that is a base object for the lepton type (*probe*-leg). We then measure the efficiency of the complete selection (*tight*) which uses the *probe* as the base object. The *tag* and *tight* selection may or may not be the same (for example we use both as TCE to assess its efficiency, but we use TCE as the tag and PHX as the tight to assess the PHX efficiency).

It is assumed that the base object efficiencies (described in Tables B.1 and B.2) are high and well modeled by the Monte Carlo. This assumption is reasonable for low statistics searches, but should probably be considered more carefully for high statistics analyses. For the forward lepton types that use tracks we have a further tracking efficiency calculation, which is applied to account for the efficiency to reconstruct the tracks in the probe for the PHX id cuts (ProbePHXPEM) and and the probe for the CMIOPES id cuts (ProbeForwardMuon).

B.1 Efficiency calculation

We use the following notation for our calculations:

- V = tagged ("very tight"),
- T = tight selection (i.e. the selection being tested), but **not tagged**,
- F = passed probe selection, but not tight.

Table B.1: Probes used for the electron efficiencies

Central Probe (TCE/LCE)
Track or SMX Fiducial to CES
Track $p_T \geq 5$
Track $\lvert Z_0 \rvert \leq 60cm$

PHX Tracking Probe
Pes2DEta $1.2 < \lvert\eta\rvert < 2$
Had/Em ≤ 0.05
Iso/Et ≤ 0.1
PEM3x3FitTower: true
$PEM3x3\chi^2 \leq 10$
Pes5x9U ≥ 0.65
Pes5x9V ≥ 0.65
ΔR(Pes,PEM) ≤ 3.0

PHX Id Probe
Pes2DEta $1.2 < \lvert\eta\rvert < 2$
Had/Em ≤ 0.05
PhxMatch: True
NSiHits ≥ 3
$\lvert TrackZ0 \rvert \leq 60cm$

PEM Probe
1.2 <PESη < 2.8
Had/Em < 0.125

Table B.2: Probes used for the muon efficiencies

Central Muon Probe (CMUP/CMX/CMIOCES/CrkTrk)
Track not PES Fiducial
NAxL(5 hits) ≥ 2
NStL(5 hits) ≥ 2
Track $\lvert Z_0 \rvert \leq 60cm$

Forward Muon Probe (CMIOPES)
Track PES Fiducial
Cot Hit Fraction > 0.6
Track $\lvert Z_0 \rvert \leq 60cm$

We can then write the efficiency of the tight selection as

$$\epsilon_{probe \to tight} = \frac{2N_{VV} + N_{VT}}{2N_{VV} + N_{VT} + N_{VF}}, \tag{B.1}$$

the N_{VV} is the number of candidates with both legs with the V condition above, N_{VT} is the number with one V and one T leg, and N_{VF} is the number with one V and one F leg. Note that if the tight and tag leg coincide then the condition T defined above can't happen, and the formula reduces to the familiar:

$$\epsilon_{probe \to tight} = \frac{2N_{VV}}{2N_{VV} + N_{VF}}. \tag{B.2}$$

Similarly if the tight and tag are mutually exclusive then N_{VV} will be zero and the formula becomes:

$$\epsilon_{probe \to tight} = \frac{N_{VT}}{N_{VT} + N_{VF}}. \tag{B.3}$$

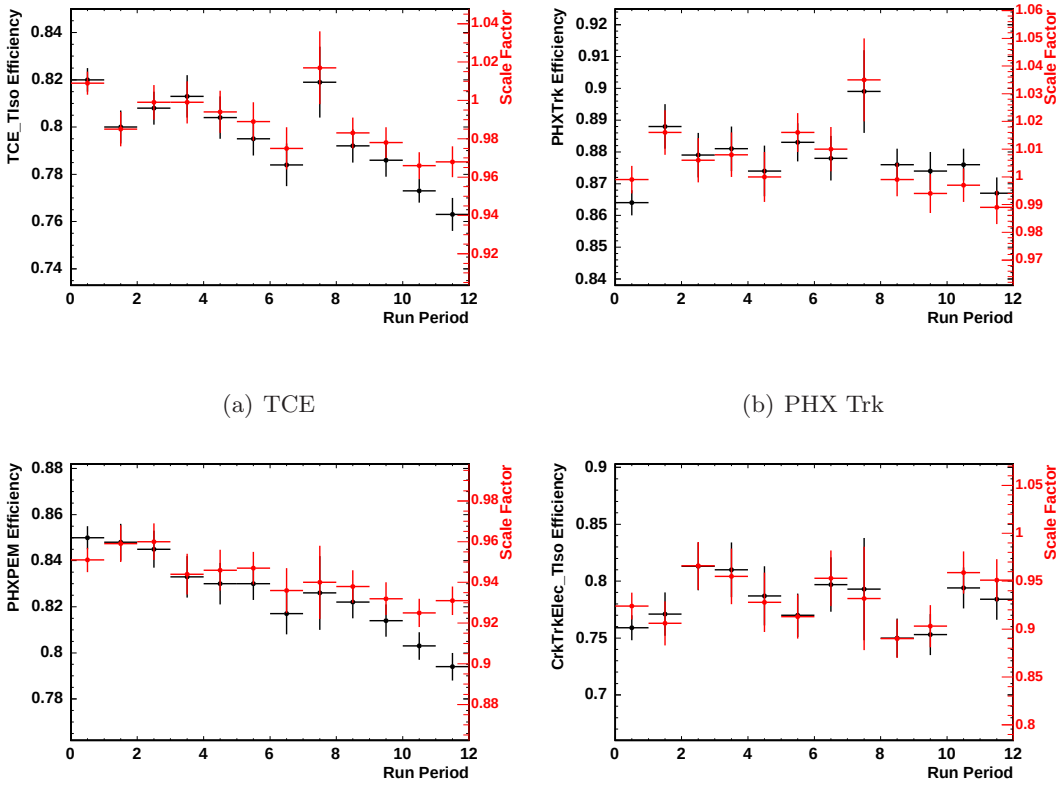
B.2 Background subtraction

We subtract the backgrounds using the Z peak sidebands. This is done for each for the N_{VV}, N_{VT}, and N_{VF} by subtracting the yields in the Z sidebands (61 to 76 GeV and 106 to 121 GeV) from the Z peak (76 to 106 GeV) yields. This procedure works regardless of the leakage l of signal into the sideband as long as the efficiency does not depend on the mass. The final yields used is:

$$\begin{aligned} N_{VV} &= \underbrace{((1-l)N_{VV}^{sig} + N_{VV}^{bkg})}_{\text{yield in peak region}} - \underbrace{(lN_{VV}^{sig} + N_{VV}^{bkg})}_{\text{yield in sideband region}} \\ &= (1-2l)N_{VV}^{sig} \end{aligned} \tag{B.4}$$

Since each of the sideband subtracted yields (N_{XX}) will be similarly scaled, the $1-2l$ factor in front will cancel in the efficiency ratio. Again this only works if l is the same regardless in which category the event falls (i.e VV, VT, or VF).

The lepton identification efficiency for each lepton category is shown in Figure B.1 B.2. The efficiency is compared between data and Monte Carlo. The 1.9 fb^{-1} data are divided into 12 run period and each period has roughly 0.16 fb^{-1} data. The efficiency decreases by 7% due to the isolation energy is larger in higher luminosity run. The overall agreement between Monte Carlo data data is within 3% and consistent with the statistical variation.



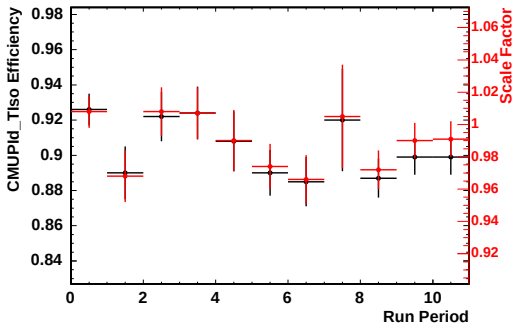
(a) TCE

(b) PHX Trk

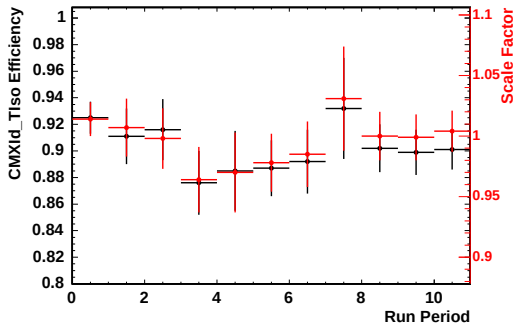
(c) PHX PEM

(d) CrkTrk Electron

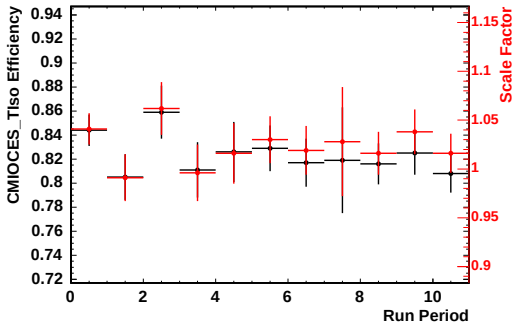
Figure B.1: Lepton identification efficiency for electrons as a function of run period.



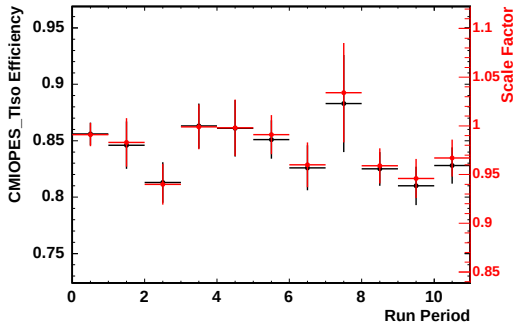
(a) CMUP



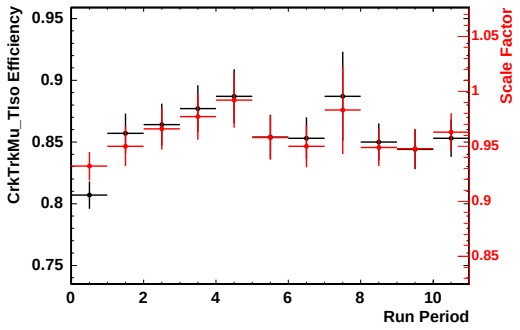
(b) CMX



(c) CMIOCES



(d) CMIOPES



(e) CrkTrk Muon

Figure B.2: Lepton identification efficiency for muons as a function of run period.

Appendix C

Fake Rate Measurement

The lepton fake rates are measured in the jet samples correcting for the small contribution of real isolated leptons from electroweak processes. This a procedure that has been done many times at CDF and elsewhere, however for each set of cuts and definitions of the denominator objects it must be repeated. The procedure is to select identified leptons (numerator objects) and denominator objects ("fakeable objects") from the jet samples, subtract the electroweak contributions and calculate the ratio

$$f_i \equiv \frac{N_i(\text{Identified Leptons}) - \sum_{j \in EWK} N_{ij}(\text{Identified Leptons})}{N_i(\text{Denominator Objects}) - \sum_{j \in EWK} N_{ij}(\text{Denominator Objects})} \qquad \text{(C.1)}$$

for each lepton category i. Using f_i, the fake lepton background predictions are $B_i = f_i \times N_i^{\text{data}}(\text{Denominator Objects})$.

The denominator definitions are described in Tables C.1 and C.2. The selection is chosen such that every tight selected lepton is a subset of its own denominator. We make a few selection cuts to avoid bias in the fake rate calculation. We do not use the leading jet to reduce the trigger bias. We require the distance from the z_0 of the lepton to the weighted mean z_0 of tracks associated with the jet to be less than 4 cm. This cut is made to simulate the effect of the Δz_0 between the leptons in multi-lepton analysis. Finally events flagged as cosmic rays are removed and the good run list appropriate to the lepton type is applied. The electroweak contributions are subtracted using the Monte Carlo samples. For each jet sample an effective luminosity is determined for the em_nosi good run list.

The resulting, f_i, fake rates are shown in Figures C.1. We use those labeled loose because they have better statistical precision, although the systematic uncertainties are slightly larger. The fake rates are averaged over the whole 1.9fb^{-1} data. We assign a

Table C.1: Denominators used for the electron fake rates. The FakeablePEM is not used in this theis.

FakeableCEM (TCE/LCE)
Fiducial Track Fiducial to CES
Had/Em $\leq 0.125 + 0.00045E$
Iso/Et ≤ 0.3
IsConversion!=1

FakeablePHX (PHX)
Pes2DEta $1.2 < \|\eta\| < 2$
Had/Em $\leq 0.125 + 0.00045E$
PhxMatch True
NSiHits ≥ 3
$\|TrackZ0\| \leq 60cm$
Iso/Et ≤ 0.3

FakeablePEM (PEM)
Pes2DEta $1.2 < \|\eta\| < 2.8$
Had/Em $\leq 0.125 + 0.00045E$
Iso/Et ≤ 0.3
No PhxMatch with
NSiHits$\geq$ 3 and $\|TrackZ0\| \leq 60cm$

Table C.2: Denominators used for the muon fake rates.

FakeableMuonCentral (CMUP/CMX/CMIOCES/CrkTrk)
Track not PES fiducial
NAxL(5 hits) ≥ 2
NStL(5 hits) ≥ 2
Track $\|Z_0\| \leq 60cm$
Track $\|D_0\| \leq 0.2cm$ (0.02 cm if NSiHit$>$ 0)
$\chi^2/dof \leq 4$ (3 if run $\leq$ 186598)
Iso/Et ≤ 0.3

FakeableMuonForward (CMIOPES)
Track PES fiducial
Cot Hit Fraction > 0.6
Track $\|Z_0\| \leq 60cm$
Track $\|D_0\| \leq 0.2cm$ (0.02 cm if NSiHit$>$ 0)
$\chi^2/dof \leq 4$ (3 if run $\leq$ 186598)
Iso/Et ≤ 0.3

systematic based on the scatter of the results between the jet samples. An uncorrelated systematic, α, is added to each jet sample measurement until all the points are consistent within 1 σ of the new mean, where the mean and the σ from the mean are calculated using $\sqrt{\sigma_{\text{stat}} \oplus \alpha}$ as the uncertainty on each point. These results are presented in the figures as a light gray band. The dark gray band is the result of a purely statistical combination.

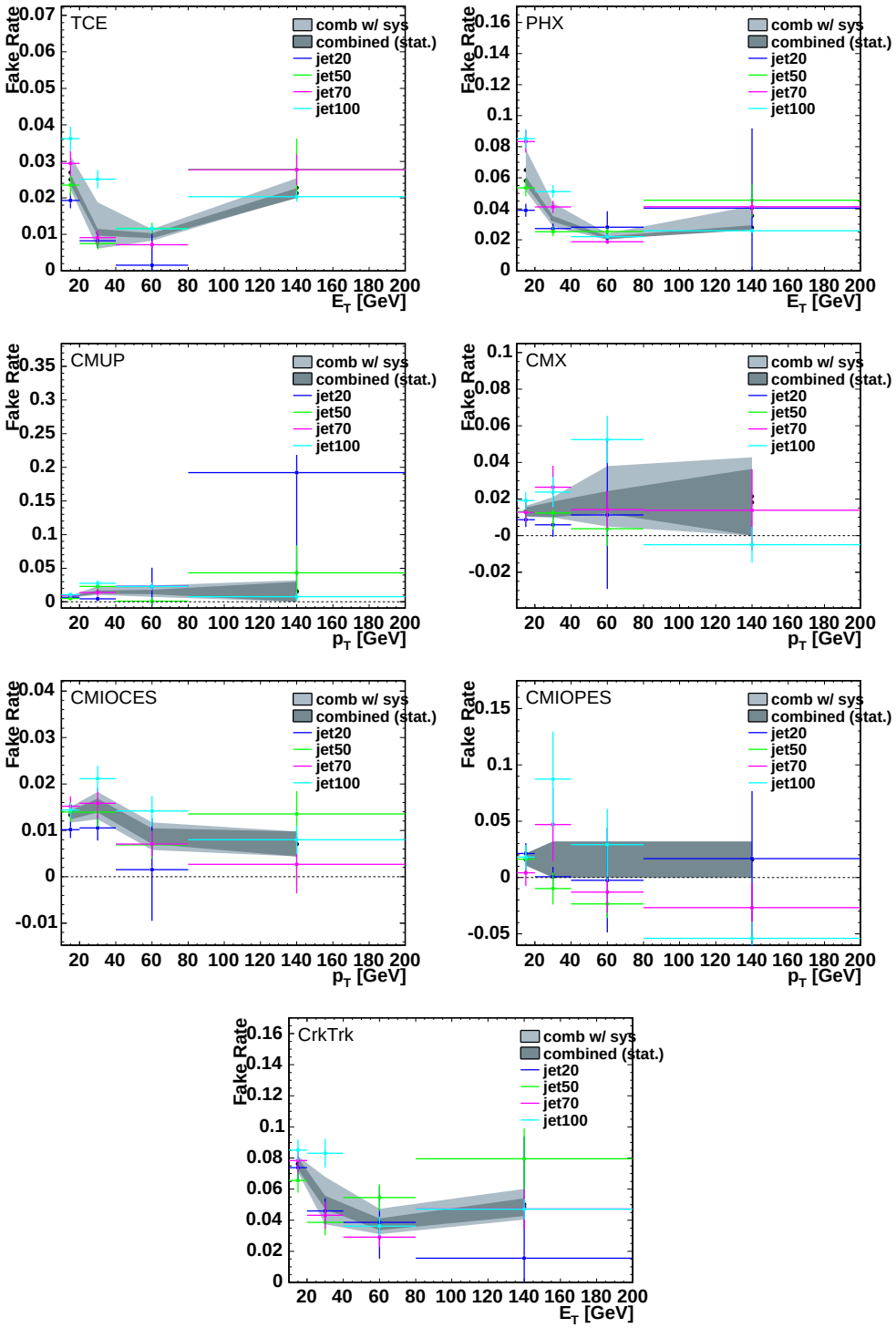


Figure C.1: Fake rate of seven types of leptons.

Appendix D

Mathematical Details of Event Probability Calculation

The details of the formula for event probability is presented in this section. A general formula with momentum balanced will be derived. To allow more efficient way of calculation, variable transformation is done and the corresponding Jacobian as well as the multiple-solutions have to be calculated appropriately.

D.1 General Formula of Event Probability

Consider measured dilepton momentum, $\vec{P_{l^+}}$ and $\vec{P_{l^-}}$, and missing transverse momentum $\not{E}_{TX}$, $\not{E}_{TY}$, the general formula for event probability is

$$P(x,\alpha) = \frac{1}{<\sigma>}\int \frac{d\sigma(y,\alpha)}{dy}\epsilon(y)G(x,y)dy \tag{D.1}$$

, where $x = (\vec{P_{l1}}, \vec{P_{l2}}, \not{E}_{Tx}, \not{E}_{Ty})$ is eight dimension vector, $y = (\vec{L^+}, \vec{L^-}, \not{\epsilon}_{Tx}, \not{\epsilon}_{Ty})$ is true value, $\epsilon(y)$ is acceptance efficiency function and $G(x,y)$ is transfer function to encounter for the resolution effects. The parton-level differential cross-section is

$$\begin{aligned}\frac{d\sigma(y,\alpha)}{dy} &= \int \frac{\sum_{ij} f_i(x_1,Q^2) f_j(x_2,Q^2)|M_{ij}(\alpha)|^2}{4\sqrt{(q_1\cdot q_2)^2 - m_{q_1}^2 m_{q_2}^2}} \\ &\quad dx_1 dx_2 d\Phi_N(z)\delta(y-z)k_T(k_x,k_y)dk_x dk_y,\end{aligned} \tag{D.2}$$

, where $z = (\vec{p_{l^+}}, \vec{p_{l^-}}, \not{p}_{Tx}, \not{p}_{Ty})$ is dummy variable have to be integrated out. This formalism has factor out the resolution and acceptance efficiency function from event

probability calculation. K_T describes the transverse momentum distribution function of the total system, $k_{(x,y)} = (\vec{p_{l^+}} + \vec{p_{l^-}} + \not{p}_T)_{(x,y)}$ is the transverse momentum of the total system where $\not{p}_T$ is the true missing energy. The leading order differential cross-section by definition has zero transverse momentum. The k_T modeling is extracted from PYTHIA to emulate the soft and hard QCD effects in terms of a boost of the total system in the transverse direction of beam axis. x_1 and x_2 are the energy fraction of two incident partons and p and q are the four momentums of two incident partons. In the case of zero transverse momentum of total system, the four vector of two incident partons could be written as

$$p = (0, 0, x_1 E_{beam}, x_1 E_{beam}), q = (0, 0, -x_2 E_{beam}, x_2 E_{beam}), \tag{D.3}$$

where $E_{beam} = 980$ GeV is the beam energy of (anti)-proton beam in Tevatron. The center-of-mass energy of the total system is $(p+q)^2 = 4x_1 x_2 E_{beam}^2$. With extra transverse momentum (k_x, k_y), the four vector of total system could be written as

$$p + q = (k_x, k_y, (x_1 - x_2)E_{beam}, \sqrt{(x_1 + x_2)^2 E_{beam}^2 - k_x^2 - k_y^2}) \tag{D.4}$$

The advantage of this convention is that the center of mass energy will still depend only on the product of x_1 and x_2 after using the property that incident partons are massless

$$(p + q)^2 = p^2 + q^2 + 2(p \cdot q) = 2(p \cdot q) = 4x_1 x_2 E_{beam}^2 \tag{D.5}$$

The flux term could be written as

$$flux = \frac{1}{4\sqrt{(p \cdot q)^2 - m_p^2 m_q^2}} = \frac{1}{8x_1 x_2 E_{beam}^2} \tag{D.6}$$

The function $f_i(x_1, Q^2)$ in $d\sigma/dy$ is the parton distribution function with factorization scale Q^2. $d\Phi_N$ is N body decay phase space as defined below

$$d\Phi_N = (2\pi)^4 \delta^4 (p + q - \sum_{i=1}^{N} p_i) \prod_{i=1}^{N} \frac{d^3 p_i}{(2\pi)^3 2E_i}, \tag{D.7}$$

where p_i is the four momentum of ith daughter particle. The δ function in the phase space calculation preserves the four momentum conservation of total system. The event probability could be written as:

$$\begin{aligned} P(x, \alpha) &= \frac{1}{<\sigma>} \int dx_1 dx_2 \frac{\sum_{ij} f_i(x_1, Q^2) f_j(x_2, Q^2) |M(\alpha)|^2}{8x_1 x_2 E_{beam}^2} \\ &\times (2\pi)^4 \delta^4 (p + q - \sum p_i) \prod \frac{d^3 p_i}{(2\pi)^3 2E_i} \\ &\times \delta^3 (L^+ - p_1) \delta^3 (L^- - p_2) \delta^2 (\not{\epsilon}_T - \not{p}_T) K_T(k_x, k_y) dk_x dk_y \\ &\times G(\vec{P_{l^+}}, \vec{P_{l^-}}, \vec{\not{E}_T}; \vec{L^+}, \vec{L^-}, \vec{\not{\epsilon}_T}) d^3 L^+ d^3 L^- d\not{\epsilon}_{Tx} d\not{\epsilon}_{Ty} \end{aligned} \tag{D.8}$$

This formula could be simplified by integrating over these delta function. The integration over x_1 and x_2 could eliminate two degrees of freedom of total momentum conservation delta function

$$\begin{aligned} & \int dx_1 dx_2 \delta(E_p + E_q - \sum E_i)\delta(p_z + q_z - \sum p_{zi}) \\ = & \int dx_1 dx_2 \\ \times & \ \delta(\sqrt{(x_1+x_2)^2 E_{beam}^2 - k_x^2 - k_y^2} - \sum E_i) \\ \times & \ \delta((x_1 - x_2)E_{beam} - \sum p_{Zi}) \\ = & \int df_1 df_2 \delta(f_1)\delta(f_2)\frac{1}{|J|} = \frac{1}{|J|}, \end{aligned} \tag{D.9}$$

where $f_1 = E_p + E_q - \sum E_i$, $f_2 = p_z + q_z - \sum p_{zi}$ and the Jocobian $|J|$ is defined as

$$|J| = |\frac{\partial(f_1, f_2)}{\partial(x_1, x_2)}| = |\begin{matrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{matrix}| = \frac{2(x_1+x_2)E_{beam}^3}{\sqrt{(x_1+x_2)^2 E_{beam}^2 - k_x^2 - k_y^2}} \sim 2E_{beam}^2, \tag{D.10}$$

x_1 and x_2 have a solution to make $f_1 = 0$ and $f_2 = 0$:

$$\begin{aligned} x_1 &= \frac{\sqrt{k_x^2 + k_y^2 + (\sum E_i)^2} + \sum p_{zi}}{2E_{beam}} \sim \frac{\sum E_i + \sum p_{zi}}{2E_{beam}} \\ x_2 &= \frac{\sqrt{k_x^2 + k_y^2 + (\sum E_i)^2} - \sum p_{zi}}{2E_{beam}} \sim \frac{\sum E_i - \sum p_{zi}}{2E_{beam}}. \end{aligned} \tag{D.11}$$

The approximation value above are evaluated at the limit $(x_1+x_2)^2 E_{beam}^2 >> k_x^2 + k_y^2$. After further integrating $\not{\epsilon}_{Tx}$ and $\not{\epsilon}_{Ty}$ over two dummy delta functions, the parton-level differential cross-section could be written as:

$$\begin{aligned} P(x,\alpha) &= \frac{1}{<\sigma>} \int \frac{\sum_{ij} f_i(x_1,Q^2) f_j(x_2,Q^2) |M(\alpha)|^2}{16 x_1 x_2 E_{beam}^4} (2\pi)^4 \prod \frac{d^3 p_i}{(2\pi)^3 2E_i} \\ &\times \ \delta(k_x - \sum p_i x)\delta(k_y - \sum p_{iy})\delta^3(L^+ - p_1)\delta^3(L^- - p_2) \\ &\times \ \epsilon(\vec{L^+}, \vec{L^-}, \vec{\not{p}_T}) G(\vec{P_{l^+}}, \vec{P_{l^-}}, \vec{\not{E}_T}; \vec{L^+}, \vec{L^-}, \vec{\not{p}_T}) K_T(k_x, k_y) \\ &\times \ d^3 L^+ d^3 L^- dk_x dk_y \\ &= \frac{1}{<\sigma>} \int \frac{d\sigma_0(y;\alpha)}{dy} \epsilon(y) G(x;y) K_T(\vec{k_T}) dy, \end{aligned} \tag{D.12}$$

where $dy = d^3 L^+ d^3 L^- K_T dk_x dk_y$ is true value for dilepton three vectors and total transverse momentum. Note that the relation of true transverse missing energy, $\not{p}_T$, has been related to the total transverse energy through $\vec{k_T} = \vec{p_{1T}} + \vec{p_{2T}} + \not{p}_T$. The event probability is re-written in a way to decompose transverse momentum integration from the parton-level differential cross-section. The advantage is that we can just focus on the parton-level differential cross-section when evaluating different physics process.

D.2 Differential Cross-section

D.2.1 Higgs I

Let's calculate the differential cross-section for Higgs with a given true dilepton vectors $\vec{L^+}$, $\vec{L^-}$ and the total transverse momentum $\vec{k_T}$.

$$\begin{aligned}\frac{d\sigma_0(y;m_H)}{dy} &= \int \frac{\sum_{ij} f_i(x_1,Q^2) f_j(x_2,Q^2)|M(m_H)|^2}{16 x_1 x_2 E_{beam}^4} \frac{(2\pi)^4}{(2\pi)^{3\times 4} 2^4} \frac{d^3 l^+}{l_E^+} \frac{d^3 l^-}{l_E^-} \frac{d^3\nu}{\nu_E} \frac{d^3\bar{\nu}}{\bar{\nu}_E} \\ &\times \delta(k_x - l_x^+ - l_x^- - \nu_x - \bar{\nu}_x)\delta(k_y - l_y^+ - l_y^- - \nu_y - \bar{\nu}_y) \\ &\times \delta^3(L^+ - p_1)\delta^3(L^- - p_2) \\ &= \int \frac{\sum_{ij} f_i(x_1,Q^2) f_j(x_2,Q^2)|M(m_H)|^2}{16(2\pi^3)^8 16 x_1 x_2 E_{beam}^4} \frac{\nu_x \nu_y \nu_z \bar{\nu}_z}{L_E^+ L_E^- \nu_E \bar{\nu}_E}.\end{aligned} \tag{D.13}$$

This is done by integrating $\vec{l^+}$, $\vec{l^-}$, $\bar{\nu}_x$ and $\bar{\nu}_y$ over delta functions. Using Monte Carlo integration approach to evaluate numerical value could be done by generating four dimensional phase space

$$\nu_x, \nu_y, \nu_z, \bar{\nu}_z. \tag{D.14}$$

The four vector of total system is completely determined. However, this is not a good choice to do integration because of the narrow Breit-Wigner width of Higgs such that the sampling of phase space is out side of the core function. A natural choice of integration phase space is Higgs mass, M_H^2, and it could be done by doing a variable transformation $\bar{\nu}_z \to M_H^2$ through

$$d\bar{\nu}_z = \frac{1}{|J|} dM_H^2 = \frac{1}{|\frac{\partial M_H^2}{\partial \bar{\nu}_z}|} dM_H^2, \tag{D.15}$$

where $\bar{\nu}_z$ could be solved through equality

$$\begin{aligned} M_H^2 &= H_E^2 - H_x^2 - H_y^2 - H_z^2 = H_E^2 - k_x^2 - k_y^2 - H_z^2 \\ &= (L^+ + L^- + \nu + \bar{\nu})^2 \\ &= (L_E^+ + L_E^- + \nu_E + \sqrt{\bar{\nu}_x^2 + \bar{\nu}_y^2 + \bar{\nu}_z^2})^2 \\ &\quad -k_x^2 - k_y^2 - (L_z^+ + L_z^- + \nu_z + \bar{\nu}_z)^2 \\ &= E^2 + 2E\sqrt{T + \bar{\nu}_z^2} + T + \bar{\nu}_z^2 \\ &\quad -k_x^2 - k_y^2 - Z^2 - 2Z\bar{\nu}_z - \bar{\nu}_z^2, \end{aligned} \tag{D.16}$$

where the parameters are re-grouped into $E = L_E^+ + L_E^- + \nu_E$, $T = \bar{\nu}_x^2 + \bar{\nu}_y^2$, $Z = L_z^+ + L_z^- + \nu_z$. There are two solutions for $\bar{\nu}_z$

$$\bar{\nu_z} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \tag{D.17}$$

where $a = 4Z^2 - 4E^2$,$b = 4ZH$, $H = M_H^2 + k_x^2 + k_y^2 + Z^2 - E^2 - T$ and c=$H^2 - 4E^2T$. The Jocobian is

$$|J| = 2H_E\frac{\bar{\nu}_z}{\bar{\nu}_E} - 2H_z \tag{D.18}$$

The formula for Higgs equation is then,

$$\frac{d\sigma_H}{dx} = \frac{1}{(2\pi)^8 16}\int\sum_m \frac{f(x_1)f(x_2)|M_H|^2}{16x_1x_2E_{Beam}^4}\frac{dM_H^2 d\nu_x d\nu_y d\nu_z}{L_E^+L_E^-\nu_E\bar{\nu}_E}\frac{1}{|J_m|}. \tag{D.19}$$

We have to sum up two solutions of $\bar{\nu}_z$ for a given M_H^2 phase space.

D.2.2 Higgs II

When Higgs mass is greater than two W mass, both W tend to be on-shell such that we can transform two neutrino integration into Higgs and W mass integration through

$$d\nu_x d\nu_y = \frac{1}{|\frac{\partial(M_H^2, M_{W^+}^2)}{\partial(\nu_x,\nu_y)}|}dM_H^2 dM_{W^+}^2 \tag{D.20}$$

ν_x and ν_y should satisfy the equality

$$\begin{aligned} M_H^2 &= (L_E^+ + L_E^- + \nu_E + \bar{\nu_E})^2 - k_x^2 - k_y^2 - (L_z^+ + L_z^- + \nu_z + \bar{\nu_z})^2 \\ M_{W^+}^2 &= (L^+ + \nu)^2 = 2(L^+\cdot\nu) = 2(L_E^+\nu_E - L_x^+\nu_x - L_y^+\nu_y - L_z^+\nu_z) \end{aligned} \tag{D.21}$$

The formula is:

$$\frac{d\sigma_H}{dx} = \frac{1}{(2\pi)^8 16}\int\frac{f(x_1)f(x_2)|M_H|^2}{16x_1x_2E_{Beam}^4}\frac{dM_H dM_W d\nu_z d\bar{\nu}_z}{L_E^+L_E^-\nu_E\bar{\nu}_E}\frac{1}{|J|}, \tag{D.22}$$

where Jacobian J is:

$$J = 4(W_E\frac{\nu_Z}{\nu_E} - W_Z)(H_E\frac{\bar{\nu}_Z}{\bar{\nu}_E} - H_Z). \tag{D.23}$$

There are four solutions:

D.2.3 WW

$$\frac{d\sigma_{WW}(y)}{dy} = \int\frac{\sum_{ij} f_i(x_1,Q^2)f_j(x_2,Q^2)|M(m_H)|^2}{16(2\pi^3)^8 16x_1x_2E_{beam}^4}\frac{\nu_x\nu_y\nu_z\bar{\nu}_z}{L_E^+L_E^-\nu_E\bar{\nu}_E}. \tag{D.24}$$

In the case of WW mass, both W are on-shell such that we could transform to two W mass.

$$d\nu_x d\nu_y = \frac{1}{|\frac{\partial(M_H^2, M_{W^+}^2)}{\partial(\nu_x,\nu_y)}|} dM_{W^+}^2 dM_{W^-}^2 \tag{D.25}$$

ν_x and ν_y should satisfy the equality

$$\begin{aligned} M_{W^+}^2 &= (L^+ + \nu)^2 = 2(L^+ \cdot \nu) = 2(L_E^+\nu_E - L_x^+\nu_x - L_y^+\nu_y - L_z^+\nu_z) \\ M_{W^-}^2 &= (L^- + \bar{\nu})^2 = 2(L^- \cdot \bar{\nu}) = 2(L_E^-\bar{\nu}_E - L_x^-\bar{\nu}_x - L_y^-\bar{\nu}_y - L_z^-\bar{\nu}_z) \end{aligned} \tag{D.26}$$

It could be written as a bi-quartic equation:

$$p_1\nu_x^2 + p_2\nu_x\nu_y + p_3\nu_y^2 + p_4\nu_x + p_5\nu_y + p_6 = 0 \tag{D.27}$$

$$q_1\nu_x^2 + q_2\nu_x\nu_y + q_3\nu_y^2 + q_4\nu_x + q_5\nu_y + q_6 = 0, \tag{D.28}$$

where

$$\begin{aligned} p_1 &= L_x^{+2} - L_E^{+2} & q_1 &= L_x^{-2} - L_E^{-2} \\ p_2 &= 2L_x^+L_y^+ & q_2 &= 2L_x^-L_y^- \\ p_3 &= L_y^{+2} - L_E^{+2} & q_3 &= L_y^{-2} - L_E^{-2} \\ p_4 &= L_x^+A & q_4 &= -2L_x^-D + 2Q_xL_E^{-2} \\ p_5 &= L_y^+A & q_5 &= -2L_y^-D + 2Q_yL_E^{-2} \\ p_6 &= \tfrac{A^2}{4} - (L_E^+\nu_z)^2 & q_6 &= D^2 - L_E^{-2}(Q_x^2 + Q_y^2 + \bar{\nu}_z^2) \\ A &= M_{W^+}^2 + 2L_Z^+\nu_Z & D &= M_{W^-}^2/2 + L_x^-Q_x + L_y^-Q_y + L_z^-\bar{\nu}_z \end{aligned}$$

Take D.27 $\times q_3$ - D.28 $\times p_3$, we can solve

$$\nu_y = \tfrac{q_3(p_1\nu_x^2+p_4\nu_x+p_6)-p_3(q_1\nu_x^2+q_4\nu_x+q_6)}{p_3(q_2\nu_x+q_5)-q_3(p_2\nu_x+p_5)} = \tfrac{M\nu_x^2+N\nu_x^2+O}{P\nu_x+Q} \tag{D.29}$$

, where

$$\begin{aligned} M &= q_3p_1 - p_3q_1 \\ N &= q_3p_4 - p_3q_4 \\ O &= q_3p_6 - p_3q_6 \\ P &= -q_3p_2 + p_3q_2 \\ Q &= -q_3p_5 + p_3q_5 \end{aligned}$$

Substitute it back to D.27. We can write down a fourth order polynomial

$$k_4\nu_x^4 + k_3\nu_x^3 + k_2\nu_x^2 + k_1\nu_x + k_0 = 0 \tag{D.30}$$

, where

$$\begin{array}{lllllll} k_4 = & & +P^2p_1 & +MR & & \\ k_3 = & 2PQp_1 & +P^2p_4 & +MS & +NR & \\ k_2 = Q^2p_1 & +2PQp_4 & +P^2p_6 & +MT & +NS & +QR \\ k_1 = Q^2p_4 & +2PQp_6 & & & +NT & +OS \\ k_0 = Q^2p_6 & & & & & +OT, \end{array}$$

where $R = Pp_2 + Mp_3$, $S = Pp_5 + Qp_2 + Np_3$ and $T = Qp_5 + Op_4$. We can use numerical method to find the roots of this equation. Using Sturm Sequences [82] to Bracket Real Roots of Polynomial Equations. There are four solutions and the final differential cross-section is:

$$\frac{d\sigma_{WW}}{dx} = \frac{1}{(2\pi)^8 16}\int \frac{f(x_1)f(x_2)|M_{WW}|^2}{16x_1x_2E_{Beam}^4}\frac{dM_{W1}dM_{W2}d\nu_x d\nu_y}{L_E^+L_E^-\nu_E\bar{\nu}_E}\frac{1}{|J|}, \tag{D.31}$$

where Jacobian J is:

$$J = 4(L_E^+\frac{\nu_y}{\nu_E} - L_y^+)(L_E^-\frac{\bar{\nu}_x}{\nu_E} - L_x^-) - 4(L_E^+\frac{\nu_x}{\nu_E} - L_x^+)(L_E^-\frac{\bar{\nu}_y}{\nu_E} - L_y^-) \tag{D.32}$$

D.2.4 ZZ

The ZZ decay channel is also four body decay.

$$\frac{d\sigma_{ZZ}(y)}{dy} = \int \frac{\sum_{ij} f_i(x_1,Q^2)f_j(x_2,Q^2)|M_{ZZ}|^2}{16(2\pi^3)^8 16x_1x_2E_{beam}^4}\frac{\nu_x\nu_y\nu_z\bar{\nu}_z}{L_E^+L_E^-\nu_E\bar{\nu}_E}. \tag{D.33}$$

Since neutrino pairs decay from Z boson and neutrino not couple to γ^* such that it's always on-shell Z. The trivial option for the integration is to transform into M_Z^2

$$\begin{array}{rcl} M_Z^2 & = & (\nu+\bar{\nu})^2 = 2\nu\cdot\bar{\nu} \\ & = & 2(\nu_E\sqrt{\bar{\nu}_x^2+\bar{\nu}_y^2+\bar{\nu}_z^2} - \nu_x\bar{\nu_x} - \nu_y\bar{\nu_y} - \nu_z\bar{\nu_z}) \end{array} \tag{D.34}$$

There are two solutions for $\bar{\nu}_z$:

$$\bar{\nu}_Z = \frac{-b \pm \sqrt{b^2-4ac}}{2a}, \tag{D.35}$$

where $a = \nu_Z^2 - \nu_E^2$, $b = 2\nu_z B$, $B = M_Z/2 + \nu_x\bar{\nu_x} + \nu_y\bar{\nu_y}$ and $c = B^2 - \nu_E^2(\bar{\nu}_x^2 + \bar{\nu}_y^2)$

$$\frac{d\sigma_{ZZ}}{dx} = \frac{1}{(2\pi)^8 16}\int \frac{f(x_1)f(x_2)|M_{ZZ}|^2}{16x_1x_2E_{Beam}^4}\frac{d\nu_x d\nu_y d\nu_z dM_Z^2}{L_{1E}L_{2E}\nu_E\bar{\nu}_E}\frac{1}{2L_{1E}n\bar{u}_Z/L_{2E} - 2\nu_Z} \tag{D.36}$$

D.2.5 $W\gamma$/jet

$$\begin{aligned}\frac{d\sigma_{W+\gamma}(y)}{dy} &= \int \frac{\sum_{ij} f_i(x_1,Q^2) f_j(x_2,Q^2)|M_{W+\gamma}|^2}{16 x_1 x_2 E_{beam}^4} \frac{(2\pi)^4}{(2\pi)^{3\times 3} 2^3} \frac{d^3 l^+}{l_E^+} \frac{d^3\gamma}{\gamma_E} \frac{d^3\nu}{\nu_E} \\ &\times \delta(k_x - l_x^+ - \gamma_x - \nu_x)\delta(k_y - l_y^+ - \gamma - \nu_y) \\ &\times \delta^3(L^+ - l^+)\delta^3(L^- - \gamma) \\ &= \int \frac{\sum_{ij} f_i(x_1,Q^2) f_j(x_2,Q^2)|M(m_H)|^2}{16(2\pi^3)^8 16 x_1 x_2 E_{beam}^4} \frac{\nu_z}{L_E^+ \gamma_E \nu_E}.\end{aligned} \tag{D.37}$$

$$\frac{d\sigma_{W^\pm\gamma}}{dx} = \frac{1}{(2\pi)^5 8} \int \frac{f(x_1) f(x_2)|M_{W^\pm\gamma}|^2}{16 x_1 x_2 E_{Beam}^2} \frac{d\nu_z}{L_E^\pm \gamma_E \nu_E} Conversion(\gamma \to L^\mp) \tag{D.38}$$

D.3 Normalization Constant

The normalization constant $< \sigma >$ for each processes are listed below. It is the expected yields for each process normalized to LO cross-section predictions.

Table D.1 shows the normalization constants used for each event probability calculation. The normalization is break down to 22 categories. The normalization constant is calculated from the expected yields of fully simulated Monte Carlo. The acceptance is calculated and normalized it to the leading order cross-section for each process. The event probability calculation is using leading order Matrix Element such that the normalization constant is also renormalized to leading order.

The other way to calculate normalization constant is to integrate over each event probability $P(x_{obs})$ over 8 dimensional observation space

$$N_{accepted} = \int P(x_{obs}) A(x_{obs}) dx_{obs}, \tag{D.39}$$

where $A(x_{obs})$ is the geometry acceptance function. Note that we have factor acceptance function from event probability for simplification of calculation in chapter 6. In principle, the normalization rescaled from fully simulated Monte Carlo will be identical to fully event probability simulation. However, the simplification of our computing makes them not be identical. An approximate normalization will only decrease our sensitivity but not bias our results once we do the template fitting to set statistics limits.

D.4 Parton Fake Rate

Figure D.1 shows the probability for a parton to be reconstructed a denominator. They are averaged results from various PYTHIA dijet Monte Carlos.

Table D.1: The acceptance table for Matrix Element. The definition of each symbol is explained as follow: T-TCE, P-PHX, C-CMUP, X-CMX, CO-CMIOCES, CP-CMIOPES, K-CrkTrk.

Type	T-T	T-P	P-P	T-C	T-X	T-OC	T-OP	P-C	P-X	P-OC	P-OP	C-C	C-X	C-OC	C-OP	X-X	X-OP	X-OC	T-K	P-K	C-K	X-K
HWW130	4.2e-02	2.2e-02	4.7e-03	2.6e-02	1.0e-02	1.6e-02	8.3e-03	6.6e-03	2.9e-03	3.2e-03	2.7e-03	6.1e-03	1.0e-02	1.5e-02	7.8e-03	1.7e-03	5.9e-03	3.2e-03	1.3e-02	2.7e-03	7.0e-03	2.7e-03
HWW140	5.0e-02	2.8e-02	5.3e-03	3.2e-02	1.2e-02	2.1e-02	1.1e-02	7.9e-03	3.4e-03	3.9e-03	3.2e-03	7.6e-03	1.3e-02	2.0e-02	9.7e-03	2.0e-03	7.6e-03	4.0e-03	1.7e-02	3.5e-03	9.0e-03	3.5e-03
HWW150	5.9e-02	3.2e-02	5.8e-03	3.6e-02	1.3e-02	2.4e-02	1.2e-02	8.7e-03	3.9e-03	4.6e-03	3.5e-03	8.7e-03	1.5e-02	2.4e-02	1.1e-02	2.4e-03	8.6e-03	4.7e-03	2.1e-02	4.1e-03	1.1e-02	4.3e-03
HWW160	6.7e-02	3.3e-02	6.7e-03	4.0e-02	1.5e-02	2.7e-02	1.3e-02	8.9e-03	4.3e-03	5.3e-03	4.0e-03	9.8e-03	1.7e-02	2.8e-02	1.2e-02	2.5e-03	9.8e-03	5.3e-03	2.4e-02	4.8e-03	1.3e-02	4.8e-03
HWW170	7.4e-02	3.7e-02	6.9e-03	4.4e-02	1.6e-02	3.0e-02	1.5e-02	9.6e-03	4.5e-03	5.7e-03	4.4e-03	1.1e-02	1.8e-02	3.1e-02	1.3e-02	2.8e-03	1.1e-02	5.9e-03	2.7e-02	5.4e-03	1.4e-02	5.5e-03
HWW180	7.8e-02	4.0e-02	7.1e-03	4.5e-02	1.7e-02	3.1e-02	1.6e-02	1.1e-02	4.5e-03	6.1e-03	4.5e-03	1.1e-02	1.8e-02	3.2e-02	1.5e-02	3.0e-03	1.1e-02	6.1e-03	2.9e-02	5.9e-03	1.5e-02	5.8e-03
HWW190	8.1e-02	4.3e-02	6.7e-03	4.7e-02	1.8e-02	3.2e-02	1.7e-02	1.2e-02	5.0e-03	7.0e-03	4.2e-03	1.2e-02	1.9e-02	3.3e-02	1.6e-02	3.2e-03	1.2e-02	6.4e-03	3.0e-02	6.5e-03	1.6e-02	6.0e-03
HWW200	8.4e-02	4.5e-02	6.6e-03	4.9e-02	1.9e-02	3.4e-02	1.8e-02	1.3e-02	4.9e-03	7.3e-03	4.4e-03	1.2e-02	2.0e-02	3.2e-02	1.7e-02	3.6e-03	1.3e-02	6.9e-03	3.1e-02	6.7e-03	1.6e-02	6.2e-03
HWW230	8.4e-02	4.5e-02	6.6e-03	4.9e-02	1.9e-02	3.4e-02	1.8e-02	1.3e-02	4.9e-03	7.3e-03	4.4e-03	1.2e-02	2.0e-02	3.2e-02	1.7e-02	3.6e-03	1.3e-02	6.9e-03	3.1e-02	6.7e-03	1.6e-02	6.2e-03
WW	5.9e-02	4.0e-02	7.0e-03	3.7e-02	1.6e-02	2.7e-02	1.6e-02	1.3e-02	5.5e-03	8.6e-03	4.9e-03	7.5e-03	1.3e-02	2.2e-02	1.3e-02	2.7e-03	9.7e-03	5.5e-03	2.3e-02	7.1e-03	1.1e-02	4.7e-03
ZZ	9.7e-02	6.2e-02	1.0e-02	0.0e-00	0.0e-00	0.0e-00	0.0e-00	0.0e-00	0.0e-00	0.0e-00	0.0e-00	0.0e-00	0.0e-00	2.2e-02	3.8e-02	2.0e-02	4.5e-03	1.6e-02	8.6e-03	2.7e-02	7.6e-03	3.2e-02	1.4e-02
Wp_gamma	3.6e-05	1.4e-04	3.8e-05	8.8e-06	8.1e-06	4.4e-06	4.9e-07	3.1e-05	1.8e-05	1.0e-05	9.7e-06	0.0e-00	0.0e-00	0.0e-00	0.0e-00	0.0e-00	0.0e-00	0.0e-00	8.8e-06	1.6e-05	4.5e-06	1.0e-06
Wm_gamma	3.6e-05	1.4e-04	3.8e-05	8.8e-06	8.1e-06	4.4e-06	4.9e-07	3.1e-05	1.8e-05	1.0e-05	9.7e-06	0.0e-00	0.0e-00	0.0e-00	0.0e-00	0.0e-00	0.0e-00	0.0e-00	8.8e-06	1.6e-05	4.5e-06	1.0e-06
Wp_1jet	5.8e-06	1.9e-05	8.5e-06	3.7e-06	1.7e-06	9.0e-07	1.8e-06	6.9e-06	3.2e-06	4.4e-07	9.0e-07	5.0e-07	1.2e-06	7.1e-07	1.4e-06	1.3e-07	2.4e-07	5.3e-07	3.5e-06	2.8e-06	1.9e-06	1.5e-06
Wm_1jet	5.8e-06	1.9e-05	8.5e-06	3.7e-06	1.7e-06	9.0e-07	1.8e-06	6.9e-06	3.2e-06	4.4e-07	9.0e-07	5.0e-07	1.2e-06	7.1e-07	1.4e-06	1.3e-07	2.4e-07	5.3e-07	3.5e-06	2.8e-06	1.9e-06	1.5e-06
WpZ_lostW	5.9e-02	4.0e-02	7.0e-03	3.7e-02	1.6e-02	2.7e-02	1.6e-02	1.3e-02	5.5e-03	8.6e-03	4.9e-03	7.5e-03	1.3e-02	2.2e-02	1.3e-02	2.7e-03	9.7e-03	5.5e-03	2.3e-02	7.1e-03	1.1e-02	4.7e-03
WpZ_lostZ	5.9e-02	4.0e-02	7.0e-03	3.7e-02	1.6e-02	2.7e-02	1.6e-02	1.3e-02	5.5e-03	8.6e-03	4.9e-03	7.5e-03	1.3e-02	2.2e-02	1.3e-02	2.7e-03	9.7e-03	5.5e-03	2.3e-02	7.1e-03	1.1e-02	4.7e-03
Z_2l	5.9e-02	4.0e-02	7.0e-03	3.7e-02	1.6e-02	2.7e-02	1.6e-02	1.3e-02	5.5e-03	8.6e-03	4.9e-03	7.5e-03	1.3e-02	2.2e-02	1.3e-02	2.7e-03	9.7e-03	5.5e-03	2.3e-02	7.1e-03	1.1e-02	4.7e-03

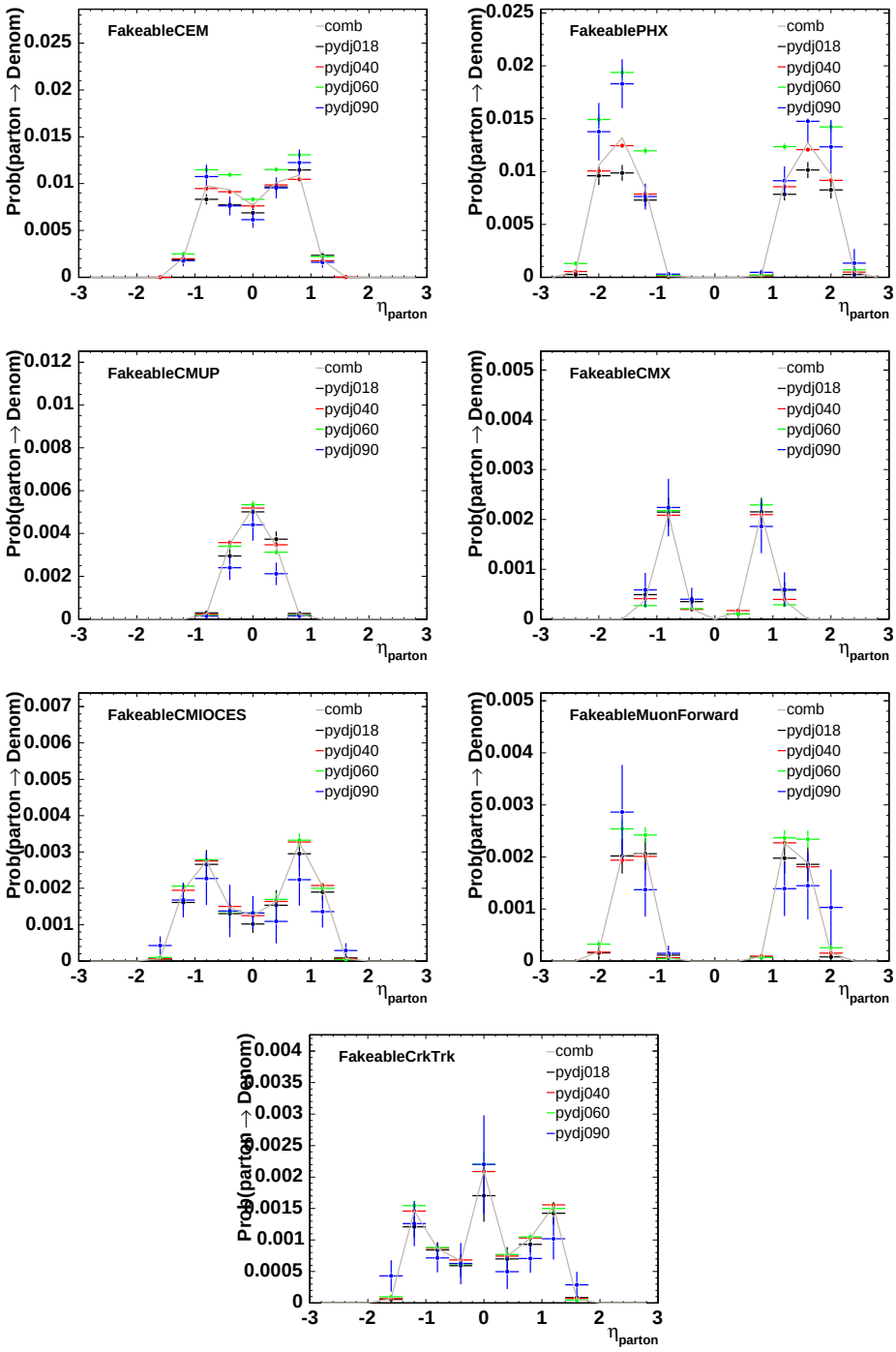


Figure D.1: The probability of a parton to be reconstructed as a denominator. The results are averaged from four PYTHIA samples with different $\hat{p_T}$, 18, 40, 60 and 90 GeV/c.

Appendix E

Miscellaneous Figures

E.1 Kinematics in the Base region

Figure E.1- E.2 show the kinematics distributions of input variables for event probability calculation at 1.9 fb^{-1}. Both Confidence Level (CL) and Kolmogorov-Smirnov (KS) Test are performed in each distribution. We only do shape comparison but no absolute rate included.

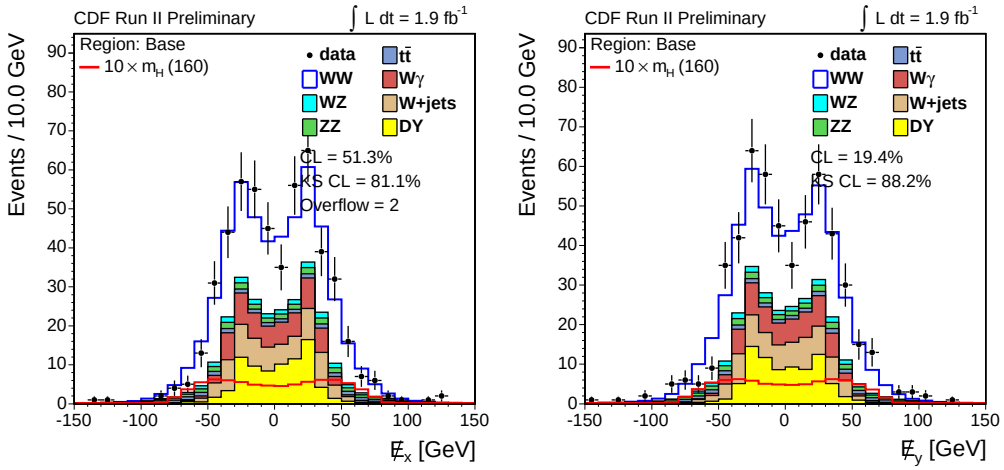


Figure E.1: Distribution of $\not\!\!E_T$ in x-y dimensions in the Base region with 1.9 fb^{-1}.

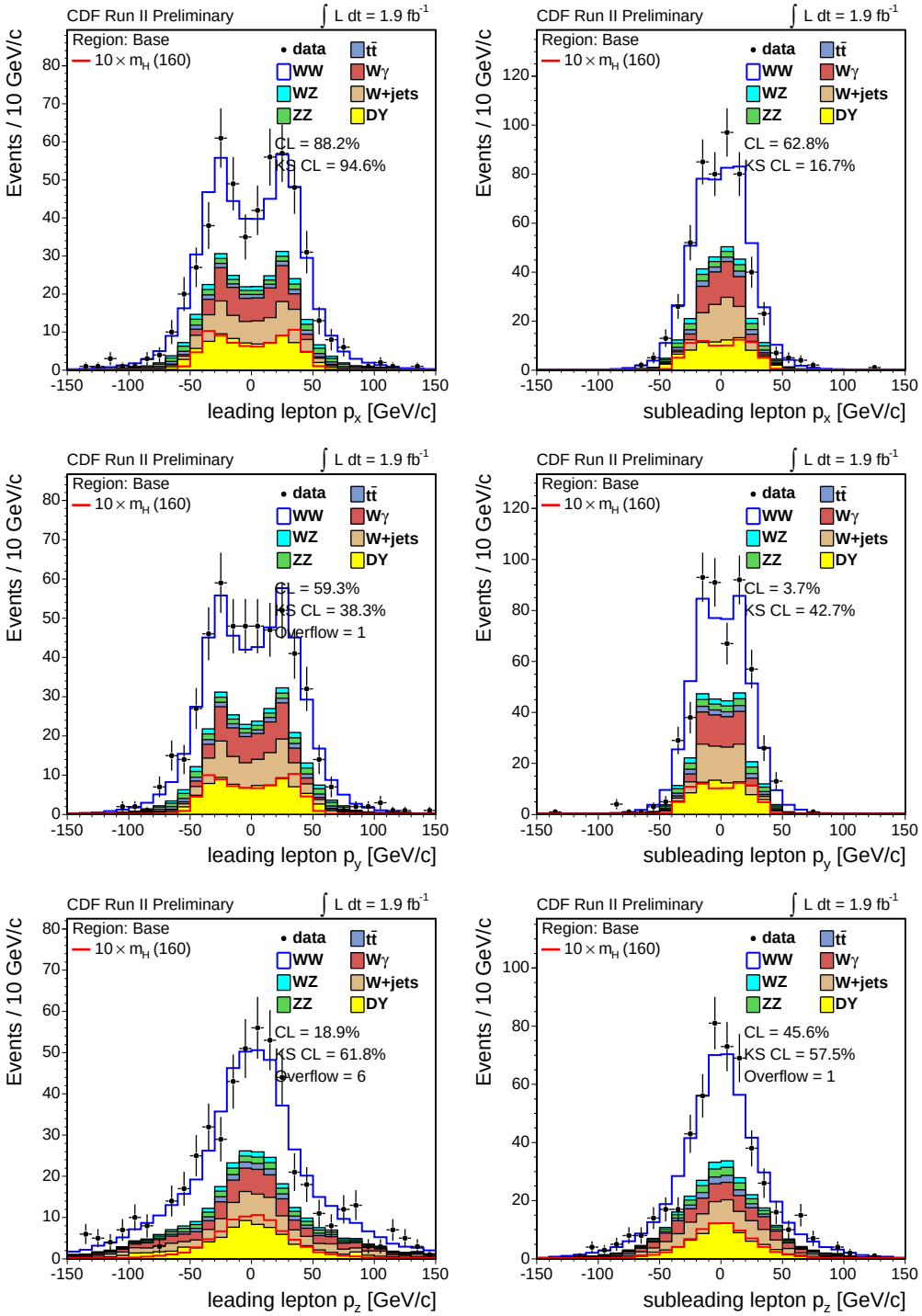
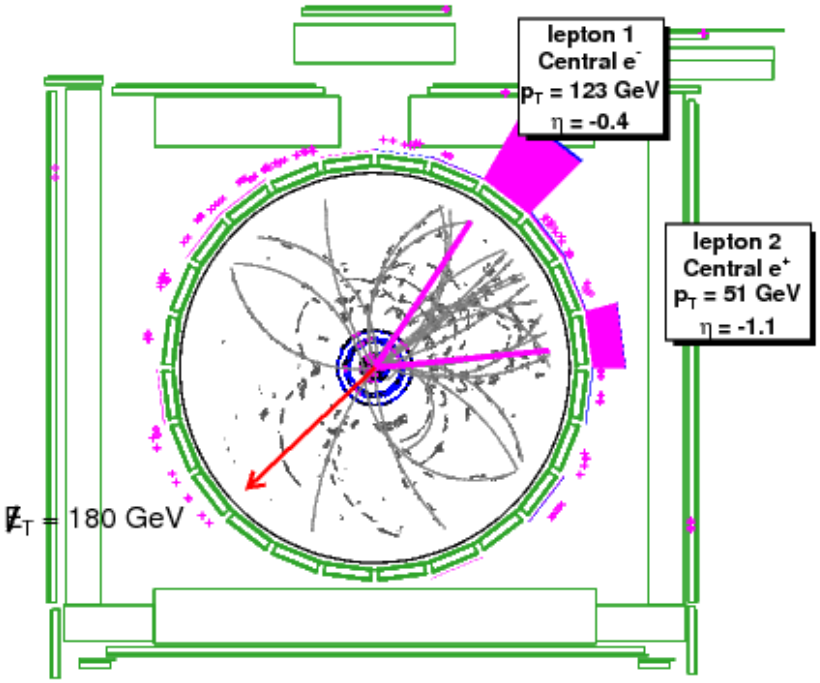


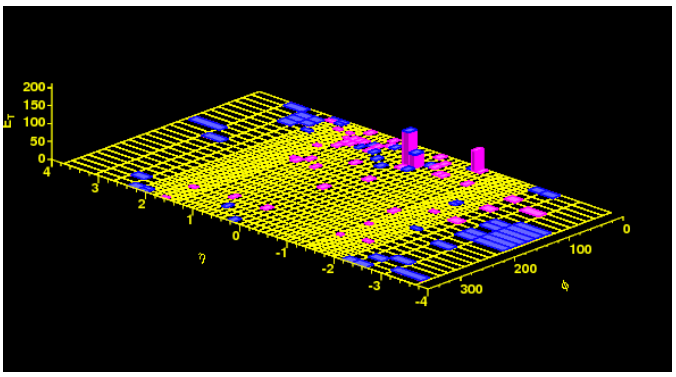
Figure E.2: Lepton momentum distribution in 3 dimension in base selection at 1.9 fb^{-1}.

E.2 Event Display of most ZZ-like Events

The event display of the most ZZ-like events are shown in Figure E.3, Figure E.4, Figure E.5.

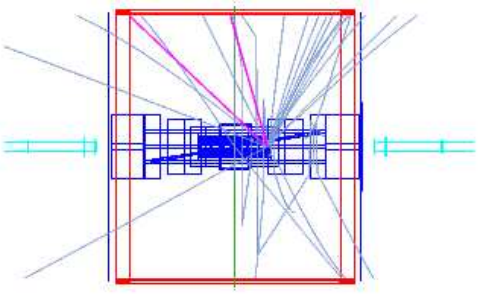


(a) $r - \phi$



(b) $\eta - \phi$

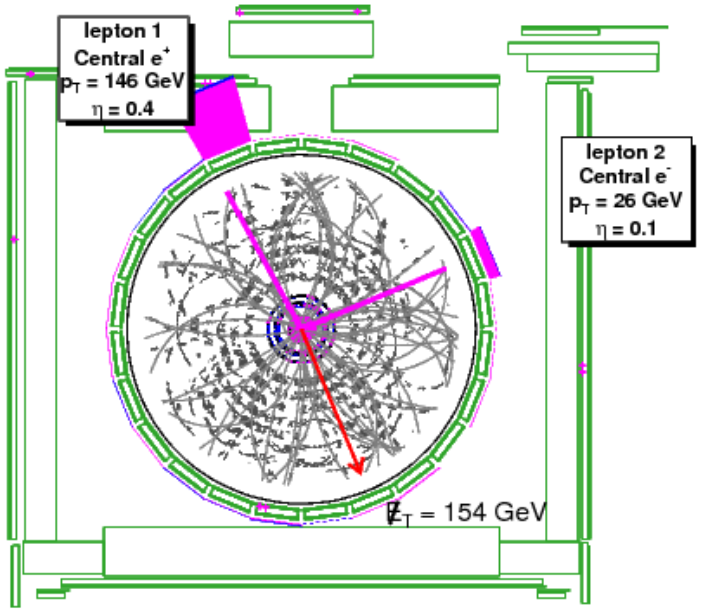
m_{ll}=91.22 GeV $|\not\!\!E_T|$=180.5 GeV



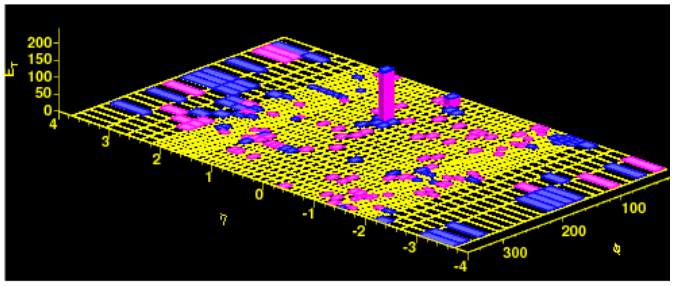
(c) $r - z$

Type	P_T	η	ϕ
Central e	123.5	-0.4	1.0
Central e	51.5	-1.1	0.1

Figure E.3: ZZ Signal Event: Run=203265 Event=3792931

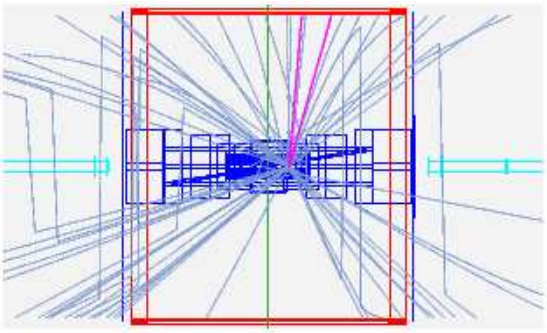


(a) $r - \phi$



(b) $\eta - \phi$

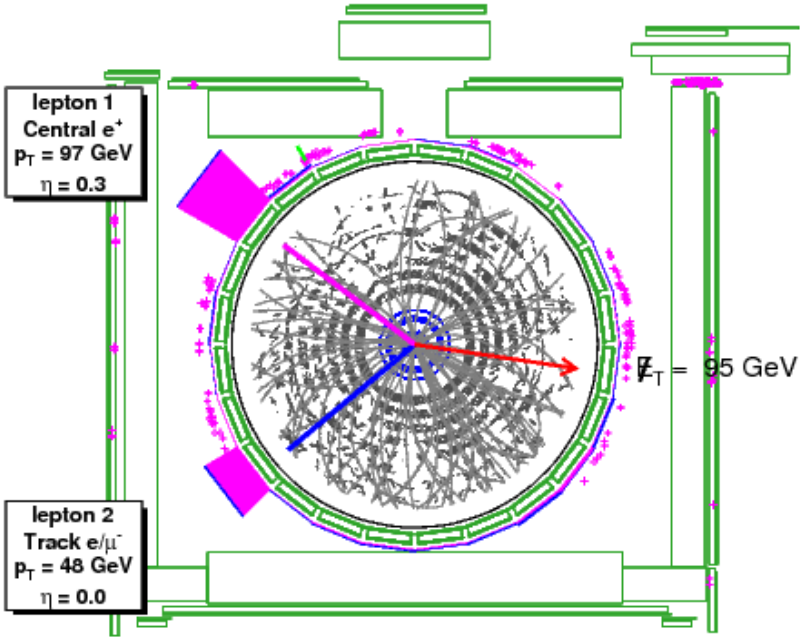
m_{ll}=94.24 GeV $|\not\!\!E_T|$=153.8 GeV



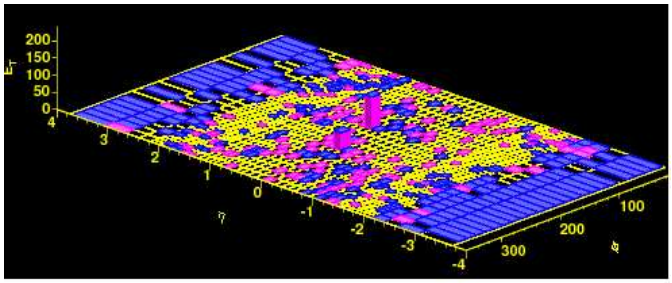
(c) $r - z$

Type	P_T	η	ϕ
Central e	146.3	0.4	2.1
Central e	26.4	0.1	0.4

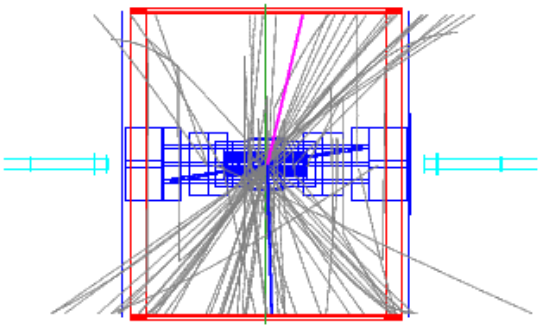
Figure E.4: ZZ Signal Event: Run=209021 Event=4038281



(a) $r - \phi$



(b) $\eta - \phi$



(c) $r - z$

m_{ll}=85.96 GeV $|\not\!\!E_T|$=94.9 GeV

Type	P_T	η	ϕ
Central e	97.0	0.3	2.5
Track e/μ	47.5	0.0	-2.5

Figure E.5: ZZ Signal Event: Run=197288 Event=11891

Bibliography

[1] W. M. Yao *et al.* (Particle Data Group), J. Phys. **G33**, 1 (2006).

[2] P. De Jong (L3) Prepared for 9th Annual Divisional Meeting (DPF 96) of the Division of Particles and Fields of the American Physical Society, Minneapolis, Minnesota, 11-15 Aug 1996.

[3] M. Acciarri *et al.* (L3), Phys. Lett. **B450**, 281 (1999).

[4] A. Abulencia *et al.* (CDF), Phys. Rev. Lett. **97**, 081802 (2006), `hep-ex/0605124`.

[5] D. E. Acosta *et al.* (CDF), Phys. Rev. Lett. **94**, 211801 (2005), `hep-ex/0501050`.

[6] D. E. Acosta *et al.* (CDF), Phys. Rev. **D71**, 091105 (2005), `hep-ex/0501021`.

[7] (2007), `0712.2383`.

[8] T. Aaltonen *et al.* (CDF), Phys. Rev. Lett. **100**, 201801 (2008), `0801.4806`.

[9] A. Abulencia *et al.* (CDF), Phys. Rev. Lett. **98**, 161801 (2007), `hep-ex/0702027`.

[10] T. Aaltonen *et al.* (CDF) (2008), `0809.3930`.

[11] S. F. Novaes (1999), `hep-ph/0001283`.

[12] M. Halzen, *Quarks and leptons* (1984), quarks and Leptons: An Introductory Course in Modern Paricle Physics.

[13] H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).

[14] S. L. Glashow, Nucl. Phys. **22**, 579 (1961).

[15] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).

[16] D. J. Gross and F. Wilczek, Phys. Rev. **D8**, 3633 (1973).

[17] S. Peskin, *An introduction to quantum field theory* (1995).

[18] A. Djouadi, J. Kalinowski, and M. Spira, Comput. Phys. Commun. **108**, 56 (1998), `hep-ph/9704448`.

[19] D. Rainwater, TASI 2006 lectures (2007), `hep-ph/0702124`.

[20] T. Hambye and K. Riesselmann, Phys. Rev. **D55**, 7255 (1997), hep-ph/9610272.

[21] J. A. Casas, J. R. Espinosa, and M. Quiros, Phys. Lett. **B382**, 374 (1996), hep-ph/9603227.

[22] R. Barate *et al.* (LEP Working Group for Higgs boson searches), Phys. Lett. **B565**, 61 (2003), hep-ex/0306033.

[23] D. C. Kennedy and B. W. Lynn, Nucl. Phys. **B322**, 1 (1989).

[24] J. Alcaraz *et al.* (LEP) (2007), 0712.0929.

[25] G. Bernardi *et al.* (D0), ICHEP-2006 conference (2006), hep-ex/0612044.

[26] V. M. Abazov *et al.* (D0), Phys. Rev. Lett. **96**, 011801 (2006), hep-ex/0508054.

[27] E. Arik, O. Cakir, S. A. Cetin, and S. Sultansoy, Acta Phys. Polon. **B37**, 2839 (2006), hep-ph/0502050.

[28] E. Arik, O. Cakir, S. A. Cetin, and S. Sultansoy, Phys. Rev. **D66**, 033003 (2002), hep-ph/0203257.

[29] W. Loinaz and J. D. Wells, Phys. Lett. **B445**, 178 (1998), hep-ph/9808287.

[30] A. V. Manohar and M. B. Wise, Phys. Lett. **B636**, 107 (2006), hep-ph/0601212.

[31] J.-L. K. G. Gounaris and D. Z. (conv.) *et al.*, CERN Report 96–01 p. p. 525 (1996).

[32] H. Aihara *et al.* (1995), hep-ph/9503425.

[33] J. M. Campbell, J. W. Huston, and W. J. Stirling, Rept. Prog. Phys. **70**, 89 (2007), hep-ph/0611148.

[34] H. L. Lai *et al.* (CTEQ), Eur. Phys. J. **C12**, 375 (2000), hep-ph/9903282.

[35] F. Corriveau (H1 and ZEUS), AIP Conf. Proc. **957**, 205 (2007).

[36] R. D. Field (CDF) (2002), hep-ph/0201192.

[37] T. Sjostrand, L. Lonnblad, and S. Mrenna (2001), hep-ph/0108264.

[38] G. Aubrecht *et al.*, Contemporary Physics Education Project (2003), http://www.lbl.gov/abc/wallchart/teachersguide/pdf/Chap11.pdf.

[39] C. W. Schmidt Presented at 1993 Particle Accelerator Conference (PAC 93), Washington, DC, 17-20 May 1993.

[40] Ferrmilab Beam Division URL http://www-bd.fnal.gov/runII/index.html.

[41] T. Affolder *et al.* (CDF), Nucl. Instrum. Meth. **A526**, 249 (2004).

[42] L. Balka *et al.* (CDF), Nucl. Instrum. Meth. **A267**, 272 (1988).

[43] S. Bertolucci *et al.* (CDF), Nucl. Instrum. Meth. **A267**, 301 (1988).

[44] F. Abe *et al.* (CDF), Phys. Rev. **D48**, 2998 (1993).

[45] A. Bhatti *et al.*, Nucl. Instrum. Meth. **A566**, 375 (2006), hep-ex/0510047.

[46] M. G. Albrow *et al.* (CDF), Nucl. Instrum. Meth. **A480**, 524 (2002).

[47] M. Aalste *et al.* (RD5), Z. Phys. **C60**, 1 (1993).

[48] T. C. I. Collaboration, FERMILAB-Pub-96/390-E (1996), http://www-cdf.fnal.gov/upgrades/tdr/tdr.html.

[49] G. Ascoli *et al.*, Nucl. Instrum. Meth. **A268**, 33 (1988).

[50] D. Acosta *et al.* (CDF), Nucl. Instrum. Meth. **A461**, 540 (2001), Private talk to Song-Ming Wang. Bunch crossing frequency should be 396 ns.

[51] D. Acosta *et al.*, Nucl. Instrum. Meth. **A494**, 57 (2002).

[52] F. Abe *et al.* (CDF), Phys. Rev. **D50**, 5550 (1994).

[53] R. Downing *et al.* (CDF), Nucl. Instrum. Meth. **A570**, 36 (2007), physics/0606247.

[54] E. J. Thomson *et al.*, IEEE Trans. Nucl. Sci. **49**, 1063 (2002).

[55] M. S. Neubauer (CDF), Nucl. Instrum. Meth. **A502**, 386 (2003).

[56] A. Kreymer *et al.* Prepared for Computing in High-Energy Physics (CHEP '04), Interlaken, Switzerland, 27 Sep - 1 Oct 2004.

[57] D. Thain, T. Tannenbaum, and M. Livny, Concurrency - Practice and Experience **17**(2-4), 323 (2005).

[58] C. Issever (CDF), AIP Conf. Proc. **670**, 371 (2003), hep-ex/0301002.

[59] U. Baur, T. Han, and J. Ohnemus, Phys. Rev. **D48**, 5140 (1993), hep-ph/9305314.

[60] S. Frixione and B. R. Webber, JHEP **06**, 029 (2002), hep-ph/0204244.

[61] The PYTHIA version 6.216 is based on 6.214 with an important bug fix implemented at Fermilab, URL http://cdfcodebrowser.fnal.gov/CdfCode/source/pythia_i/doc/pythia.html.

[62] R. Brun, R. Hagelberg, M. Hansroul, and J. C. Lassalle CERN-DD-78-2-REV.

[63] G. Grindhammer, M. Rudowicz, and S. Peters, Nucl. Instrum. Meth. **A290**, 469 (1990).

[64] S. Catani, D. de Florian, M. Grazzini, and P. Nason, JHEP **07**, 028 (2003), hep-ph/0306211.

[65] J. M. Campbell and R. K. Ellis, Phys. Rev. **D60**, 113006 (1999), hep-ph/9905386.

[66] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason, and G. Ridolfi, JHEP **04**, 068 (2004), hep-ph/0303085.

[67] B. A. Jayatilaka UMI-32-37982, Fermilab-Thesis-2006-17.

[68] B. Stelzer (CDF), Conf. Proc. of 42th Rencontres de Moriond EW 2007 (2007), 0706.0282.

[69] *VEGAS Monte Carlo Integration in GSL - GNU Scientific Library*, http://www.gnu.org/software/gsl/.

[70] J. Campbell and K. Ellis, *MCFM - Monte Carlo for FeMtobarn Processes* (2005), URL http://mcfm.fnal.gov/.

[71] C. Anastasiou, K. Melnikov, and F. Petriello, Nucl. Phys. **B724**, 197 (2005), hep-ph/0501130.

[72] CDF Parameters, http://www-cdf.fnal.gov/internal/detectors/parameters.html.

[73] T. Junk, *Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds, and Large Systematics* (2006), CDF/DOC/STATISTICS/PUBLIC/8128, http://www-cdf.fnal.gov/people/links/ThomasJunk/.

[74] C. Anastasiou, K. Melnikov, and F. Petriello, Nucl. Phys. **B724**, 197 (2005), hep-ph/0501130.

[75] A. J. Attal UMI-32-52363.

[76] J. Pumplin *et al.*, JHEP **07**, 012 (2002), hep-ph/0201195.

[77] S.-C. Hsu (CDF) (2007), 0706.2200.

[78] G. Bernardi *et al.* (Tevatron New Phenomena Higgs Working Group) (2008), 0808.0534.

[79] L. Chakravarti and Roy (1967), handbook of Methods of Applied Statistics, Volume I, John Wiley and Sons, pp. 392-394.

[80] P. J. Dervan, A. Signer, W. J. Stirling, and A. Werthenbach, J. Phys. **G26**, 607 (2000), hep-ph/0002175.

[81] B. Han and E. Halkiadakis, *Met pem trigger efficiency for pheonix electrons* (2006), CDF/DOC/ELECTRON/CDFR/7940.

[82] D. Hook and P. McAree, *Using Sturm Sequences to Bracket Real Roots of Polynomial Equations* (1990), Graphics Gems, URL http://tog.acm.org/GraphicsGems/gems/Sturm/.